INVESTOR IN PEOPLE

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

EPO - DG 1

02 03. 2001

(78)

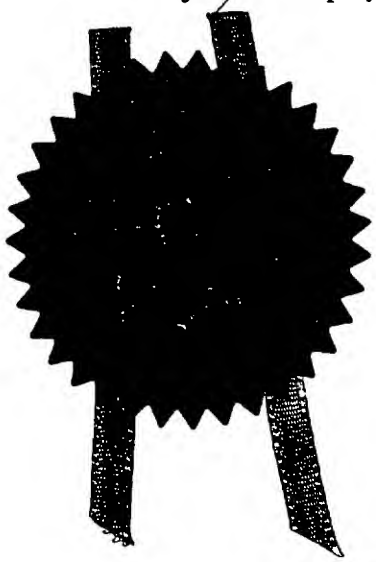
I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

I also certify that the attached copy of the request for grant of a Patent (Form 1/77) bears an amendment, effected by this office, following a request by the applicant and agreed to by the Comptroller-General.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.



Signed

Dated 22 February 2001

Patents Form 1/77

Patents Act 1977
(Rule 16)

The Patent Office

THE PATENT OFFICE
E
14 AUG 1998
LONDON

1/77

9
17AUG98 E383546-2 D02882
P01/7700 25.00 - 9817796.7

The Patent Office

Cardiff Road
Newport
Gwent NP9 1RH

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)

1.	Your reference	SCB/50899/000
2.	Patent application number *(The Patent Office will fill in this part)*	14 AUG 1998 9817796.7
3.	Full name, address and postcode of the or of each applicant *(underline all surnames)*	JANSSEN PHARMACEUTICA N.V. Turnhoutseweg 30 B-2340 Beerse BELGIUM
	Patents ADP number *(if you know it)*	
	If the applicant is a corporate body, give the country/state of its incorporation	BELGIUM 005 31939001
4.	Title of the invention	DRUG TARGETS IN CANDIDA ALBICANS
5.	Name of your agent *(if you have one)*	BOULT WADE TENNANT
	"Address for service" in the United Kingdom to which all correspondence should be sent *(including the postcode)*	~~27 FURNIVAL STREET~~ Verulam Gardens ~~LONDON~~ ~~EC4A 1PQ~~ 70 Gray's Inn Road London WC1X 8BT
	Patents ADP number *(if you know it)*	42001

6.	If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and *(if you know it)* the or each application number	Country	Priority application number (if you know it)	Date of filing *(day/month/year)*

7.	If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application	Number of earlier application	Date of filing *(day / month / year)*

8.	Is a statement of inventorship and of right to grant of a patent required in support of this request? *(Answer 'Yes' if:* *a) any applicant named in part 3 is not an inventor, or* *b) there is an inventor who is not named as an applicant, or* *c) any named applicant is a corporate body.* *See note (d))*	YES

Patents Form 1/77

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form	
Description	27
Claim(s)	5
Abstract	
Drawing(s)	74

10. If you are also filing any of the following, state how many against each item.

Priority documents	
Translations of priority documents	
Statement of inventorship and right to grant of a patent *(Patents Form 7/77)*	
Request for preliminary examination and search *(Patents Form 9/77)*	
Request for substantive examination *(Patents Form 10/77)*	
Any other documents *(Please specify)*	

11. I/We request the grant of a patent on the basis of this application.

Signature Date

14 August 1998

12. Name and daytime telephone number of person to contact in the United Kingdom

COLM D. MURPHY
0171 404 5921

Warning

After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents ct 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

a) *If you need help to fill in this form or you have any questions, please contact the Patent Office on 01645 500505.*

b) *Write your answers in capital letters using black ink or you may type them.*

c) *If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.*

d) *If you have answered 'Yes' Patents Form 7/77 will need to be filed.*

e) *Once you have filled in the form you must remember to sign and date it.*

f) *For details of the fee and ways to pay please contact the Patent Office.*

DRUG TARGETS IN CANDIDA ALBICANS

The present invention is concerned with the identification of genes or functional fragments thereof from *Candida albicans* which are critical for growth and cell division and which genes may be used as selective drug targets to treat *Candida albicans* associated infections. Novel nucleic acid sequences from *Candida albicans* are also provided and which encode the polypeptides which are critical for growth of *Candida albicans*.

Opportunistic infections in immunocompromised hosts represent an increasingly common cause of mortality and morbidity. *Candida* species are among the most commonly identified fungal pathogens associated with such opportunistic infections, with *Candida albicans* being the most common species. Such fungal infections are thus problematical in, for example, AIDS populations in addition to normal healthy women where *Candida albicans* yeasts represent the most common cause of vulvovaginitis.

Although compounds do exist for treating such disorders, such as for example, amphotericin, these drugs are generally limited in their treatment because of their toxicity and side effects. Therefore, there exists a need for new compounds which may be used to treat *Candida* associated infections in addition to compounds which are selective in their action against *Candida albicans*.

Classical approaches for identifying anti-fungal compounds have relied almost exclusively on inhibition of fungal or yeast growth as an endpoint. Libraries of natural products, semi-synthetic, or synthetic chemicals are screened for their ability to kill or arrest growth of the target pathogen or a related

nonpathogenic model organism. These tests are cumbersome and provide no information about a compounds mechanism of action. The promising lead compounds that emerge from such screens must then be tested for possible host-toxicity and detailed mechanism of action studies must subsequently be conducted to identify the affected molecular target.

The present inventors have now identified a range of nucleic acid sequences from *Candida albicans* which encode polypeptides which are critical for its survival and growth. These sequences represent novel targets which can be incorporated into an assay to selectively identify compounds capable of inhibiting expression of such polypeptides and their potential use in alleviating diseases or conditions associated with *Candida albicans* infection.

Therefore, according to a first aspect of the invention there is provided a nucleic acid molecule encoding a polypeptide which is critical for survival and growth of the yeast *Candida albicans* and which nucleic acid molecule comprises any of the sequences of nucleotides illustrated in Figures 1, 2, 4 to 7, 9 to 11, 13, 15 to 20, 22 to 26, 28 to 32, 34 to 43, 45a and b, 47 to 49, 51, 52, 53 to 57, 59 and 60.

A further aspect of the invention comprises a nucleic acid molecule encoding a polypeptide which is critical for survival and growth of the yeast *Candida albicans* and which nucleic acid molecule comprises any of the sequences of nucleotides illustrated in Figure 1, 2, 36, 37a and b, 38, 39 and 40 and fragments or derivatives of said nucleic acid molecules.

Letters utilised in the sequences according to the invention which are not recognisable as letters of the genetic code signify a position in the nucleic acid sequence where one or more of bases A, G, C or T

can occupy the nucleotide position. Representative letters used to identify the range of bases which can be used are as follows:

M:	A or C
R:	A or G
W:	A or T
S:	C or G
Y:	C or T
K:	G or T
V:	A or C or G
H:	A or C or T
D:	A or G or T
B:	C or G or T
N:	G or A or T or C

In one embodiment of each of the above identified aspects of the invention the nucleic acid may comprise a mRNA molecule or alternatively a DNA and preferably a cDNA molecule.

Also provided by the present invention is a nucleic acid molecule capable of hybridising to the nucleic acid molecules illustrated in any of Figures 1 to 61 under high stringency conditions.

Stringency of hybridisation as used herein refers to conditions under which polynucleic acids are stable. The stability of hybrids is reflected in the melting temperature (Tm) of the hybrids. Tm can be approximated by the formula:

$$81.5°C+16.6(1\log_{10}[Na^+]+0.41\ (\%G\&C)-6001/l$$

wherein l is the length of the hybrids in nucleotides. Tm decreases approximately by 1-1.5°C with every 1% decrease in sequence homology.

The nucleic acid capable of hybridising to nucleic acid molecules according to the invention will generally be at least 70%, preferably at least 80 or 90% and more preferably at least 95% homologous to the nucleotide sequences illustrated in any of Figures 1 to 61.

The DNA molecules according to the invention may, advantageously, be included in a suitable expression vector to express polypeptides encoded therefrom in a suitable host.

An expression vector according to the invention includes a vector having a nucleic acid according to the invention operably linked to regulatory sequences, such as promoter regions, that are capable of effecting expression of said DNA fragments. The term "operably linked" refers to a juxta position wherein the components described are in a relationship permitting them to function in their intended manner. Such vectors may be transformed into a suitable host cell to provide for expression of a polypeptide according to the invention. Thus, in a further aspect, the invention provides a process for preparing polypeptides according to the invention which comprises cultivating a host cell, transformed or transfected with an expression vector as described above under conditions to provide for expression by the vector of a coding sequence encoding the polypeptides, and recovering the expressed polypeptides.

The vectors may be, for example, plasmid, virus or phage vectors provided with an origin of replication, optionally a promoter for the expression of said nucleotide and optionally a regulator of the promoter. The vectors may contain one or more selectable markers, such as, for example, ampicillin

resistance.

Polynucleotides according to the invention may be inserted into the vectors described in an antisense orientation in order to provide for the production of antisense RNA. Antisense RNA or other antisense nucleic acids may be produced by synthetic means.

In accordance with the present invention, a defined nucleic acid includes not only the identical nucleic acid but also any minor base variations including in particular, substitutions in bases which result in a synonymous codon (a different codon specifying the same amino acid residue) due to the degenerate code in conservative amino acid substitutions. The term "nucleic acid sequence" also includes the complementary sequence to any single stranded sequence given regarding base variations.

The present invention also comprises within its scope proteins or polypeptides expressed by the nucleic acid molecules according to the invention or a functional equivalent, derivative or bioprecursor thereof.

The present invention also advantageously provides nucleic acid sequences of at least approximately 15 contiguous nucleotides of a nucleic acid according to the invention and preferably from 15 to 50 nucleotides. These sequences may, advantageously be used as probes or primers to initiate replication, or the like. Such nucleic acid sequences may be produced according to techniques well known in the art, such as by recombinant or synthetic means. They may also be used in diagnostic kits or the like for detecting the presence of a nucleic acid according to the invention. These tests generally comprise contacting the probe with the sample under hybridising conditions and detecting for the presence

of any duplex or triplex formation between the probe and any nucleic acid in the sample.

Advantageously, the nucleic acid sequences, according to the invention may be produced using such recombinant or synthetic means, such as for example using PCR cloning mechanisms which generally involve making a pair of primers, which may be from approximately 15 to 50 nucleotides to a region of the gene which is desired to be cloned, bringing the primers into contact with mRNA, cDNA, or genomic DNA from a human cell, performing a polymerase chain reaction under conditions which bring about amplification of the desired region, isolated the amplified region or fragment and recovering the amplified DNA. Generally, such techniques as defined herein are well known in the art, such as described in Sambrook *et al* (Molecular Cloning: a Laboratory Manual, 1989).

The nucleic acids or oligonucleotides according to the invention may carry a revealing label. Suitable labels include radioisotopes such as ^{32}P or ^{39}S, enzyme labels or other protein labels such as biotin or fluorescent markers. Such labels may be added to the nucleic acids or oligonucleotides of the invention and may be detected using known techniques *per se*.

The polypeptide or protein according to the invention includes all possible amino acid variants encoded by the nucleic acid molecule according to the invention including a polypeptide encoded by said molecule and having conservative amino acid changes. Polypeptides according to the invention further include variants of such sequences, including naturally occurring allelic variants which are substantially homologous to said polypeptides. In

this context, substantial homology is regarded as a sequence which has at least 70%, preferably 80 or 90% amino acid homology with the polypeptides encoded by he nucleic acid molecules according to the invention.

Nucleic acids and polypeptides which are particularly preferred are those comprising the sequences of nucleotides illustrated in Figures 1 and 2. These sequences are specific to *Candida albicans* with no functionally related sequences in other prokaryotic or eukaryotic organism as yet identified from the respective genomic databases.

Nucleotide sequences according to the invention are particularly advantageous for selective therapeutic targets for treating *Candida albicans* associated infections. For example, an antisense nucleic acid capable of binding to the nucleic acid sequence illustrated in any of Figures 1 to 61 may be used to selectively inhibit expression of the corresponding polypeptides, leading to impaired growth of the *Candida albicans* with reductions of associated illnesses or diseases. The antisense nucleic acid corresponding to the sequences identified in Figures 1 and 2 may therefore be particularly useful in selective treatment of *Candida albicans* associated infection.

The nucleic acid molecule or the polypeptide according to the invention may be used as a medicament, or in the preparation of a medicament, for treating diseases or conditions associated with *Candida albicans* infection.

Advantageously, the nucleic acid molecule or the polypeptide according to the invention may be provided in a pharmaceutical composition together with a pharmaceutically acceptable carrier, diluent or excipient therefor.

Antibodies to the protein or polypeptide of the present invention may, advantageously, be prepared by techniques which are known in the art. For example, polyclonal antibodies may be prepared by inoculating a host animal, such as a mouse, with the polypeptide according to the invention or an epitope thereof and recovering immune serum. Monoclonal antibodies may be prepared according to known techniques such as described by Kohler R. and Milstein C., Nature (1975) 256, 495-497.

Antibodies according to the invention may also be used in a method of detecting for the presence of a polypeptide according to the invention, which method comprises reacting the antibody with a sample and identifying any protein bound to said antibody. A kit may also be provided for performing said method which comprises an antibody according to the invention and means for reacting the antibody with said sample.

Proteins which interact with the polypeptide of the invention may be identified by investigating protein-protein interactions using the two-hybrid vector system first proposed by Chien *et al* (1991).

This technique is based on functional reconstitution in vivo of a transcription factor which activates a reporter gene. More particularly the technique comprises providing an appropriate host cell with a DNA construct comprising a reporter gene under the control of a promoter regulated by a transcription factor having a DNA binding domain and an activating domain, expressing in the host cell a first hybrid DNA sequence encoding a first fusion of a fragment or all of a nucleic acid sequence according to the invention and either said DNA binding domain or said activating domain of the transcription factor, expressing in the host at least one second hybrid DNA sequence, such as

a library or the like, encoding putative binding proteins to be investigated together with the DNA binding or activating domain of the transcription factor which is not incorporated in the first fusion; detecting any binding of the proteins to be investigated with a protein according to the invention by detecting for the presence of any reporter gene product in the host cell; optionally isolating second hybrid DNA sequences encoding the binding protein.

An example of such a technique utilises the GAL4 protein in yeast. GAL4 is a transcriptional activator of galactose metabolism in yeast and has a separate domain for binding to activators upstream of the galactose metabolising genes as well as a protein binding domain. Nucleotide vectors may be constructed, one of which comprises the nucleotide residues encoding the DNA binding domain of GAL4. These binding domain residues may be fused to a known protein encoding sequence, such as for example the nucleic acids according to the invention. The other vector comprises the residues encoding the protein binding domain of GAL4. These residues are fused to residues encoding a test protein. Any interaction between polypeptides encoded by the nucleic acid according to the invention and the protein to be tested leads to transcriptional activation of a reporter molecule in a GAL-4 transcription deficient yeast cell into which the vectors have been transformed. Preferably, a reporter molecule such as ß-galactosidase is activated upon restoration of transcription of the yeast galactose metabolism genes.

Further provided by the present invention is one or more *Candida albicans* cells comprising an induced mutation in the DNA sequence encoding the polypeptide according to the invention.

A further aspect of the invention provides a method of identifying compounds which selectively inhibit expression of polypeptides expressed from the nucleotides sequences illustrated in any of Figures 1 to 61 and which are critical for growth and survival of *Candida albicans*, which method comprises (a) contacting a compound to be tested with one or more *Candida albicans* cells having a mutation in a nucleic acid molecule according to the invention which mutation results in overexpression or underexpression of said polypeptides in addition to one or more wild type *Candida* cells, (b) monitoring the growth and/or activity of said mutated cell compared to said wild type wherein differential growth or activity of said one or more mutated *Candida* cells provides an indication of selective action of said compound on said polypeptide or another polypeptide in the same or a parallel pathway.

Compounds identifiable or identified using the method according to the invention, may advantageously be used as a medicament, or in the preparation of a medicament to treat diseases or conditions associated with *Candida albicans* infection. These compounds may also advantageously be included in a pharmaceutical composition together with a pharmaceutically acceptable carrier, diluent or excipient therefor.

A further aspect of the invention provides a method of identifying DNA sequences from a cell or organism which DNA encodes polypeptides which are critical for growth or survival, which method comprises (a) preparing a cDNA or genomic library from said cell or organism in a suitable expression vector which vector is such that it can either integrate into the genome in said cell or that it permits transcription of antisense RNA from the nucleotide

sequences in said cDNA or genomic library, (b) selecting transformants exhibiting impaired growth and determining the nucleotide sequence of the cDNA or genomic sequence from the library included in the vector from said transformant. Preferably, the cell or organism may be any yeast or filamentous fungi, such as, for example, *Saccharomyces cervisiae*, *Saccharomyces pombe* or *Candida albicans*.

A further aspect of the invention provides a pharmaceutical composition comprising a compound according to the invention together with a pharmaceutically acceptable carrier, diluent or excipient therefor.

The present invention may be more clearly understood with reference to the accompanying example, which is purely exemplary, with reference to the accompanying drawings, wherein

Figures 1 & 2: are nucleotide sequences of previously unknown function isolated from *Candida albicans* and which sequences are not present in the public domain.

Figures 3 to 35: are nucleotide sequences of previously unknown function isolated from *Candida albicans* and which sequences are partially or fully present in the public domain.

Figures 36 to 40: are nucleotide sequences isolated from *Candida albicans* and which have an identified function based on sequence homology with proteins

from other organisms and which sequences are not present in the public domain.

Figures 41 to 61: are nucleotide sequences having an identified function based on sequence homology comparisons from other organisms and which sequences are fully or partially present in the public domain.

Figure 62: is a diagrammatic representation of plasmid pGAL1PNiST-1.

Figure 63: is a nucleotide sequence of plasmid pGAL1PNiST-1 of Figure 62.

Figure 64: is a diagrammatic representation of plasmid pGAL1PSiST-1.

Figure 65: is a nucleotide sequence of plasmid pGAL1PSiST-1 of Figure 64.

Figures 66 to 106: are amino acid sequences of the appropriately corresponding DNA sequences illustrated in Figures 1 to 61.

Example 1

Identification of novel drug targets in *C. albicans* by anti-sense and disruptive integration

The principle of the approach is based on the fact that when a particular *C. albicans* mRNA is inhibited by producing the complementary anti-sense

RNA, the corresponding protein will decrease. If this protein is critical for growth or survival, the cell producing the anti-sense RNA will grow more slowly or will die.

Since anti-sense inhibition occurs at mRNA level, the gene copy number is irrelevant, thus allowing applications of the strategy even in diploid organisms.

Anti-sense RNA is endogenously produced from an integrative or episomal plasmid with an inducible promoter; induction of the promoter leads to the production of a RNA encoded by the insert of the plasmid. This insert will differ from one plasmid to another in the library. The inserts will be derived from genomic DNA fragments or from cDNA to cover-to the extent possible- the entire genome.

The vector is a proprietary vector allowing integration by homologous recombination at either the homologous insert or promoter sequence in the *Candida* genome. After introducing plasmids from cDNA or genomic libraries into *C. albicans*, transformants are screened for impaired growth after promoter (& thus anti-sense) induction in the presence of lithium acetate. Lithium acetate prolongs the G1 phase and thus allows anti-sense to act during a prolonged period of time during the cell cycle. Transformants which show impaired growth in both induced and non-induced media, thus showing a growth defect due to integrative disruption, are selected as well.

Transformants showing impaired growth are supposed to contain plasmids which produce anti-sense RNA to mRNAs critical for growth or survival. Growth is monitored by measuring growth-curves over a period of time in a device (Bioscreen Analyzer, Labsystems) which allows simultaneous measurement of growth-curves

of 200 transformants.

Subsequently plasmids can be recovered from the transformants and the sequence of their inserts determined, thus revealing which mRNA they inhibit. In order to be able to recover the genomic or cDNA insert which has integrated into the *Candida* genome, genomic DNA is isolated, cut with an enzyme which cuts only once into the library vector (and estimated approx. every 4096 bp in the genome) and religated. PCR with primers flanking the insert will yield (partial) genomic or cDNA inserts as PCR fragments which can directly be sequenced. This PCR analysis (on ligation reaction) will also show us how many integrations occurred. Alternatively the ligation reaction is transformed to E. coli and PCR analysis is performed on colonies or on plasmid DNA derived thereof.

This method is employed for a genome wide search for novel *C. albicans* genes which are important for growth or survival.

Materials & Methods

Construction of pGal1PNiST-1

The backbone of the pGAL1PNiST-1 vector (integrative anti-sense *Sfi*I-*Not*I vector) is pGEM11Zf(+) (Promega Inc.). First, the CaMAL2 *Eco*RI/*Sal*I promoter fragment from pDBV50 (D.H. Brown *et al.*) was ligated into *Eco*RI/*Sal*I-opened pGEM11Zf(+) resulting in the intermediate construct pGEMMAL2P-1. Into the latter (*Msc*I/CIP) the CaURA3 selection marker was cloned as a *Eco*47III/*Xmn*I fragment derived from pRM2. The resulting pGEMMAL2P-2 vector was *Not*I/*Hin*dIII opened in order to accept the *Not*I-stuffer-*Sfi*I cassette from pPCK1NiSCYCT-1 (*Eag*I/*Hin*dIII fragment): pMAL2PNiST-1. Finally, the plasmid pGAL1PNiST-1 was constructed by exchanging the

*Sal*I/*Ecl*136II MAL2 promoter in pMAL2PNiST-1 by the *Xho*I/*Sma*I GAL1 promoter fragment derived from pRM2GAL1P.

Construction of pGal1PSiST-1

The vector pGAL1PSiST-1 was created for cloning the small genomic DNA fragments (flanked by *Sfi*I sites) behind the GAL1 promoter. The only difference with pGAL1PNiST-1 is that the hIFNβ (stuffer fragment) insert fragment in pGAL1PSiST-1 is flanked by two *Sfi*I sites in stead of a *Sfi*I and a *Not*I site as in pGAL1PNiST-1. To construct pGAL1PSiST-1 the *Eco*RI-*Hind*III fragment, containing hIFNβ flanked by a *Sfi*I and a *Not*I site, of pMAL2pHiET-3 (unpublished) was exchanged by the *Eco*RI-*Hind*III fragment, containing hIFNβ flanked by two *Sfi*I sites, from YCp50S-S (an *E. coli / S. cerevisiae* shuttle vector derived from the plasmid YCp50, which is deposited in the ATCC collection (number 37419; Thrash *et al.*, 1985); an *Eco*RI-*Hind*III fragment, containing the gene hIFNβ, which is flanked by two *Sfi*I sites, was inserted in YCp50, creating YCp50S-S), resulting into plasmid pMAL2PSiST-1. The *mal*2 promoter from pMAL2PSiST-1 (by a *Nae*I-*Fsp*I digest) was further replaced by the *gal*1 promoter from pGAL1PNiST-1 (via a *Xho*I-*Sal*I digest), creating the vector pGAL1PSiST-1.

***Candida albicans* genomic library**

* *Preparation of the genomic DNA fragments*

A *Candida albicans* genomic DNA library with small DNA fragments (400 to 1,000 bp) was prepared. Genomic DNA of *Candida albicans* B2630 was isolated following a modified protocol of Blin and Stafford (1976). The quality of the isolated genomic DNA was checked by gel electrophoresis. Undigested DNA was located on the gel

above the marker band of 26,282 bp. A little smear, caused by fragmentation of the DNA, was present. To obtain enrichment for genomic DNA fragments of the desired size, the genomic DNA was partially digested. Several restriction enzymes (*AluI*, *HaeIII* and *RsaI*; all creating blunt ends) were tried out. The appropriate digest conditions have been determined by titration of the enzyme. Enrichment of small DNA fragments was obtained with 70 units of *AluI* on 10 µg of genomic DNA for 20 min. T4 DNA polymerase (Boehringer) and dNTPs (Boehringer) were added to polish the DNA ends. After extraction with phenol-chloroform the digest was size-fractionated on an agarose gel. The genomic DNA fragments with a length of 500 to 1,250 bp were eluted from the gel by centrifugal filtration (Zhu et *al.*, 1985). *SfiI* adaptors (5' GTTGGCCTTTT) or (5' AGGCCAAC) were attached to the DNA ends (blunt) to facilitate cloning of the fragments into the vector. Therefore, a 8-mer and 11-mer oligonucleotide (comprising the *SfiI* site) were kinated and annealed. After ligation of these adaptors to the DNA fragments a second size-fractionation was performed on an agarose gel. The DNA fragments of 400 to 1150 bp were eluted from the gel by centrifugal filtration.

* *Preparation of the pGAL1PSiST-1 vector fragment*

The small genomic DNA fragments were cloned after the GAL1 promoter in the vector pGAL1PSIST-1. Qiagen-purified pGAL1PSiST-1 plasmid DNA was digested with *SfiI* and the largest vector fragment eluted from the gel by centrifugal filtration (Zhu *et al.*, 1985). Ligation with a control DNA fragment, flanked by *SfiI* sites, was performed as a control. The ligation mix was electroporated to MC1061 *E. coli* cells. Plasmid DNA of 24 clones was analyzed. In all cases the

control fragment was inserted in the pGAL1PSiST-1 vector fragment.

* *Upscaling*

All genomic DNA fragments (450 ng) were ligated into the pGAL1PSiST-1 vector (20 ng). After electroporation at 2500V, 40μF circa 400,000 clones were obtained. These clones were pooled into three groups and stored as glycerol slants. Also Qiagen-purified DNA was prepared from these clones. A clone analysis showed an average insert length of 600 bp and a percentage of 91 for clones with an insert. The size of the library corresponds to 5 times the diploid genome. The genomic DNA inserts are sense or anti-sense orientated in the vector.

Candida albicans cDNA library

Total RNA was extracted from *Candida albicans* B2630 grown on respectively minimal (SD) and rich (YPD) medium as described by Chirgwin *et al* in Sambrook *et al*. mRNA was prepared from total RNA using the Invitrogen Fast Track procedure.

First strand cDNA is synthesised with the Superscript Reverse Transcriptase (BRL) and with an oligo dT-*Not*I Primer adapter. After second strand synthesis, cDNA is polished with Klenow enzyme and purified over a Sephacryl S-400 spun column. Phosphorylated *Sfi*I adapters are then ligated to the cDNA, followed by digestion with the *Not*I restriction enzyme. The *Sfi*I/*Not*I cDNA is then purified and sized on a Biogel column A150M.

First fraction contains approximately 38,720 clones by transformation, the second fraction only 1540 clones. Clone analysis:

Fr. I: 22/24 inserts, 16 $\geq$ 1000 bp, 4 $\geq$ 2000 bp,

average size: 1500 bp.
Fr. II: 9/12 inserts, 3 ≥ 1000 bp, average size: 960 bp cDNA was ligated in a *NotI*/*SfiI* opened pGAL1PNiST-1 vector (anti-sense)

***Candida* transformation**

The host strain used for transformation is a C. *albicans* ura3 mutant, CAI-4, which contains a deletion in orotidine-5'-phosphate decarboxylase and was obtained from William Fonzi, Georgetown University (Fonzi and Irwin). CAI-4 was transformed with the above described cDNA library or genomic library using the Pichia spheroplast module (Invitrogen). Resulting transformants were plated on minimal medium supplemented with glucose (SD, 0.67% or 1.34% Yeast Nitrogen base w/o amino acids + 2% glucose) plates and incubated for 2-3 days at 30°C.

Screening for mutants

Starter cultures were set up by inoculating each colony in 1 ml SD medium and incubating overnight at 30°C and 300 rpm. Cell densities were determined using a Coulter counter (Coulter Z1; Coulter electronics limited). 250.000 cells/ml were inoculated in 1 ml SD medium and cultures were incubated for 24 hours at 30°C and 300 rpm. Cultures were washed in minimal medium without glucose (S) and the pellet resuspended in 650 µl S medium. 8 µl of this culture is used for inoculating 400 µl cultures in a Honeywell-100 plate (Bioscreen analyzer; Labsystems). Each transformant was grown during three days in S medium containing LiAc; pH 6.0, with 2% glucose/2% maltose or 2% galactose/2% maltose respectively while shaking every 3 minutes for 20 seconds. Optical densities were measured every hour during three consecutive days and

growth curves were generated (Bioscreen analyzer; Labsystems).

Growth curves of transformants grown in respectively anti-sense non-inducing (glucose/maltose) and inducing (galactose/maltose) medium are compared and those transformants showing impaired growth upon anti-sense induction are selected for further analysis. Transformants showing impaired growth by virtue of integration into a critical gene are also selected.

Isolation of genomic or cDNA inserts

Putatively interesting transformants are grown in 1.5 ml SD overnight and genomic DNA is isolated using the Nucleon MI Yeast kit (Clontech). Concentration of genomic DNA is estimated by analyzing a sample on an agarose gel.

20 ng of genomic DNA is digested for three hours with an enzyme that cuts uniquely in the library vector (SacI for the genomic library; PstI for the cDNA library) and treated with RNAse. Samples are phenol/chloroform extracted and precipitated using NaOAc/ethanol.

The resulting pellet is resuspended in 500 μl ligation mixture (1 x ligation buffer and 4 units of T4 DNA ligase; both from Boehringer) and incubated overnight at 16°C.

After denaturation (20 min 65°C), purification (phenol/chloroform extraction) and precipitation (NaOAc/ethanol) the pellet is resuspended in 10 μl MilliQ (Millipore) water.

PCR analysis

Inverse PCR is performed on 1 μl of the precipitated ligation reaction using library vector

specific primers (oligo23 5′ TGC-AGC-TCG-ACC-TCG-ACT-G 3′ and oligo25 5′ GCG-TGA-ATG-TAA-GCG-TGA-C 3′ for the genomic library; 3pGALNistPCR primer :5'TGAGCAGCTCGCCGTCGCGC 3' and 5pGALNistPCR primer: 5'GAGTTATACCCTGCAGCTCGAC 3' for the cDNA library; both from Eurogentec) for 30 cycles each consisting of (a) 1 min at 95 °C, (b) 1 min at 57 °C, and (c) 3 min at 72 °C. In the reaction mixture 2.5 units of Taq polymerase (Boehringer) with TaqStart antibody (Clontech) (1:1) were used, and the final concentrations were 0.2 μM of each primer, 3 mM MgCl2 (Perkin Elmer Cetus) and 200 μM dNTPs (Perkin Elmer Cetus). PCR was performed in a Robocycler (Stratagene).

Sequence determination

Resulting PCR products were purified using PCR purification kit (Qiagen) and were quantified by comparison of band intensity on EtBr stained agarose gel with the intensity of DNA marker bands. The amount of PCR product (expressed in ng) used in the sequencing reaction is calculated as the length of the PCR product in basepairs divided by 10. Sequencing reactions were performed using the ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kit according to the instructions of the manufacturer (PE Applied Biosystems, Foster City, CA) except for the following modifications.

The total reaction volume was reduced to 15 μl. Reaction volume of individual reagents were changed accordingly. 6.0 μl Terminator Ready Reaction Mix was replaced by a mixture of 3.0 μl Terminator Ready Reaction Mix + 3.0 μl Half Term (GENPAK Limited, Brighton, UK). After cycle sequencing, reaction mixtures were purified over Sephadex G50 columns

prepared on Multiscreen HV opaque microtiter plates (Millipore, Molsheim, Fr) and were dried in a speedVac. Reaction products were resuspended in 3 μl loading buffer. Following denaturation for 2 min at 95°C, 1 μl of sample was applied on a 5% Long Ranger Gel (36 cm well-to-read) prepared from Singel Packs according to the supplier's instructions (FMC BioProducts, Rockland, ME). Samples were run for 7 hours 2X run on a ABI 377XL DNA sequencer. Data collection version 2.0 and Sequence analysis version 3.0 (for basecalling) software packages are from PE Applied Biosystems. Resulting sequence text files were copied onto a server for further analysis.

Sequence analysis

Nucleotide sequences were imported in the VectorNTI software package (InforMax Inc, North Bethesda, MD, USA), and the vector and insert regions of the sequences were identified. Sequence similarity searches against public and commercial sequence databases were performed with the BLAST software package (Altschul et al., 1990) version 1.4. Both the original nucleotide sequence and the six-frame conceptual translations of the insert region were used as query sequences. The used public databases were the EMBL nucleotide sequence database (Stoesser et al., 1998), the SWISS-PROT protein sequence database and its supplement TrEMBL (Bairoch and Apweiler, 1998), and the ALCES *Candida albicans* sequence database (Stanford University, University of Minnesota). The commercial sequence databases used were the LifeSeq® human and PathoSeq™ microbial genomic databases (Incyte Pharmaceuticals Inc., Palo Alto, CA, USA), and the GENESEQ patent sequence database (Derwent, London, UK). Three major results were obtained on the basis of

the sequence similarity searches: function, novelty, and specificity. A putative function was deduced on the basis of the similarity with sequences with a known function, the novelty was based on the absence or presence of the sequences in public databases, and the specificity was based on the similarity with vertebrate homologues.

Methods

Blastx of the nucleic acid sequences against the appropriate protein databases: Swiss-Prot for clones of which the complete sequence is present in the public domain, and paorfp (PathoSeq™) for clones of which the complete sequences is not present in the public domain.

The protein to which the translated nucleic acid sequence corresponds to is used as a starting point. The differences between this protein and our translated nucleic acid sequences are marked with a double line and annotated above the protein sequence. The following symbols are used:

a one-letter amino acid code or the ambiguity code X is used if our translated nucleic acid sequence has another amino acid on a certain position,

the stop codon sign *is used if our translated nucleic acid sequence has a stop codon on a certain position,

The letters fs (frame shift) are used if a frame shift occurs in our translated nucleic acid sequence, and another reading frame is used,

the words ambiguity or ambiguities are used if a part of our translated nucleic acid sequence is present in the proteins, but not visible in the alignments of the blast results,

The phrase missing sequence is used if the

translated nucleic acid sequence does not comprise that part of the protein.

Blastx: compares the six-frame conceptual translation products of a nucleotide query sequence (both strands) against a protein sequence database.

Screening for compounds modulating expression of polypeptides critical for growth and survival of *C. albicans*

The method proposed is based on observations (Sandbaken *et al.*, 1990; Hinnebusch and Liebman 1991; Ribogene PCT WO 95/11969, 1995) suggesting that underexpression or overexpression of any component of a process (e.g. translation) could lead to altered sensitivity to an inhibitor of a relevant step in that process. Such an inhibitor should be more potent against a cell limited by a deficiency in the macromolecule catalyzing that step and/or less potent against a cell containing an excess of that macromolecule, as compared to the wild type (WT) cell.

Mutant yeast strains, for example, have shown that some steps of translation are sensitive to the stoichiometry of macromolecules involved. (Sandbaken *et al.*). Such strains are more sensitive to compounds which specifically perturb translation (by acting on a component that participates in translation) but are equally sensitive to compounds with other mechanisms of action.

This method thus not only provides a means to identify whether a test compound perturbs a certain process but also an indication of the site at which it exerts its effect. The component which is present in altered form or amount in a cell whose growth is affected by a test compound is potentially the site of action of the test compound.

The assay to be set up involves measurement of growth of an isogenic strain which has been modified only in a certain specific allele, relative to a wild type (WT) *C. albicans* strain, in the presence of R-compounds. Strains can be ones in which the expression of a specific essential protein is impaired upon induction of anti-sense or strains which carry disruptions in an essential gene. An in silico approach to finding novel essential genes in *C. albicans* will be performed. A number of essential genes identified in this way will be disrupted (in one allele) and the resulting strains can be used for comparative growth screening.

Assay for High Throughput screening for drugs

35 µl minimal medium (S medium + 2% galactose + 2% maltose) is transferred in a transparent flat-bottomed 96 well plate using an automated pipetting system (Multidrop, Labsystems). A 96-channel pipettor (Hydra, Robbins Scientific) transfers 2.5 µl of R-compound at 10^{-3} M in DMSO from a stock plate into the assay plate.

The selected C. *albicans* strains (mutant and parent (CAI-4) strain) are stored as glycerol stocks (15%) at -70ºC. The strains are streaked out on selective plates (SD medium) and incubated for two days at 30ºC. For the parent strain, CAI-4, the medium is always supplemented with 20 µg/ml uridine. A single colony is scooped up and resuspended in 1 ml minimal medium (S medium + 2% galactose + 2% maltose). Cells are incubated at 30ºC for 8 hours while shaking at 250 rpm. A 10 ml culture is inoculated at 250.000 cells/ml. Cultures are incubated at 30ºC for 24 hours while shaking at 250 rpm. Cells are counted in Coulter counter and the final culture (S medium + 2%

galactose + 2% maltose) is inoculated at 20.000 to 50.000 cells/ml. Cultures are grown at 30ºC while shaking at 250 rpm until a final OD of 0.24 (+/- 0.04) 6nM is reached.

200 µl of this yeast suspension is added to all wells of MW96 plates containing R-compounds in a 450 µl total volume. MW96 plates are incubated (static) at 30ºC for 48 hours.

Optical densities are measured after 48 hours.

Test growth is expressed as a percentage of positive control growth for both mutant (x) and wild type (y) strains. The ratio (x/y) of these derived variables is calculated.

References

Thrash C., Bankier A. T., Barrell B. G. and Sternglanz R. (1985) *Proc. Natl. Acad. Sci.* USA **82**: 4374-4378.

Blin N. and Stafford D. W. (1976) *Nucleic Acids Res.* **3**: 2303-2308.

Zhu J., Kempenaers W., Van Der Straeten D., Contreras R., and Fiers W. (1985) *Bio/Technology* **3**: 1014-1016.

Sambrook J., Fritsch E.F. and Maniatis T. (1989) Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Fonzi W. and Irwin H. (1993) *Genetics* **134**:717-728.
Altschul S.F, Gish W., Miller W., Myers E.W., Lipman D.J. (1990) *J. Mol. Biol.* **215**(3):403-410.

Bairoch A. and Apweiler R. (1998) *Nucleic Acids Res.* **26**(1):38-42.

Stoesser G., Moseley M.A., Sleep J., McGowran M., Garcia-Pastor M., Sterk P. (1998) *Nucleic Acids Res.* **26**(1):8-15.

Chien et *al.*, (1991) Proc. Natl. Acad. Sci USA 88, 9578-9582.

Sandbaken M.G. Lupisella J.A., DiDomenico B. and Chakraburtty K., *J. Biol. Chem.* **265**:15838-15844, 1990.

Hinnebusch A.G. and Liebman S.W., in : The Molecular Biology of the Yeast *Saccharomyces*, Broach J.R., Pringle J.R. and Jones E.W., eds., CSH Laboratory

Press; NY 1991.

Patent application RiboGene Inc., PCT WO 95./11969, 1995.

Brown, D.H. *et al.*, Molecular General Genetics 251 (1) 75-80, 1996.

Claims

1. A nucleic acid molecule encoding a polypeptide which is critical for survival and growth of the yeast *Candida albicans* and which nucleic acid molecule comprises any of the sequences of nucleotides illustrated in Figure 1, 2, 4 to 7, 9 to 11, 13, 15 to 20, 22 to 26, 28 to 32, 34 to 43, 45a and b, 47 to 49, 51, 52, 53 to 57, 59 and 60.

2. A nucleic acid molecule encoding a polypeptide which is critical for survival and growth of the yeast *Candida albicans* and which nucleic acid molecule comprises any of the sequences of nucleotides illustrated in Figure 1, 2, 36, 37a, 38, 39 and 40 and fragments or derivatives of said nucleic acid molecules

3. A nucleic acid molecule according to claim 1 or 2 which is mRNA.

4. A nucleic acid molecule according to claim 1 or 2 which is DNA.

5. A nucleic acid molecule according to claim 4 which is cDNA.

6. A nucleic acid molecule capable of hybridising to the molecules according to any of claims 1 to 5 or the sequences illustrated in any of Figures 1 to 61 under high stringency conditions.

7. A polypeptide encoded by the nucleic acid molecule according to any of claims 1 to 6 or the sequence illustrated in any of Figures 1 to 61.

8. An expression vector comprising a nucleic acid molecule according to claim 4 or 5.

9. An expression vector according to claim 8 which comprises an inducible promoter.

10. An expression vector according to claim 8 or 9 which comprises a sequence encoding a reporter molecule.

11. A nucleic acid molecule according to any of claims 1 to 6 or the nucleotide sequences illustrated in Figure 1 to 61 for use as a medicament.

12. Use of a nucleic acid molecule according to any of claims 1 to 5 or the sequences illustrated in Figure 1 to 61 in the preparation of a medicament for treating *Candida albicans* associated diseases.

13. A polypeptide according to claim 7 for use as a medicament.

14. Use of a polypeptide according to claim 7 in the preparation of a medicament for treating *Candida albicans* associated infections.

15. A pharmaceutical composition comprising a nucleic acid molecule according to any of claims 1 to 6 or a polypeptide according to claim 7 together with a pharmaceutically acceptable carrier diluent or excipient therefor.

16. A *Candida albicans* cell comprising an induced mutation in the DNA sequence encoding the polypeptide according to claim 7.

17. A method of identifying compounds which selectively modulate expression of polypeptides which are crucial for growth and survival of *Candida albicans*, which method comprises:

(a) contacting a compound to be tested with one or more *Candida albicans* cells having a mutation in a nucleic acid molecule according to any of claims 1 to 5 which mutation results in overexpression or underexpression of said polypeptides in addition to contacting one or more wild type *Candida albicans* cells with said compound,

(b) monitoring the growth and/or activity of said mutated cell compared to said wild type; wherein differential growth or activity of said one or more mutated Candida cells is indicative of selective action of said compound on a polypeptide or another polypeptide in the same or a parallel pathway.

18. A compound identifiable according to the method of claim 17.

19. A compound according to claim 18 for use as a medicament.

20. Use of a compound according to claim 18 in the preparation of a medicament for treating *Candida albicans* associated diseases.

21. A pharmaceutical composition comprising a compound according to claim 18 together with a pharmaceutically acceptable carrier, diluent or excipient therefor.

22. A method of identifying DNA sequences from a cell or organism which DNA encodes polypeptides which are critical for growth or survival of said cell or organism, which method comprises:

(a) preparing a cDNA or genomic library from said cell or organism in a suitable expression vector which vector is such that it can either integrate into the genome in said cell or that it permits transcription of antisense RNA from the nucleotide sequences in said cDNA or genomic library.

(b) selecting transformants exhibiting impaired growth and determining the nucleotide sequence of the cDNA or genomic sequence from the library included in the vector from said transformant.

23. A method according to claim 22 wherein said cell or organism is a yeast or filamentous fungi.

24. A method according to claim 22 or 23 wherein said cell or organism is any of *Saccharomyces cervisiae*, *Saccharomyces pombe* or *Candida albicans*.

25. Plasmid pGAL1PSiST-1 having the sequence of nucleotides illustrated in Figure 63.

26. Plasmid pGAL1PNiST-1 having the sequence of nucleotides illustrated in Figure 65.

27. An antibody capable of binding to a polypeptide according to claim 7.

28. An oligonucleotide comprising a fragment of from 15 to 50 contiguous nucleic acid sequences of a

nucleic acid molecule according to any of claims 1 to 6.

Sequences with unknown function, C. albicans sequence NOT present in the public domain (ALCES/EMBL)

>328c2 1803bp in-house: 1123-1803 public: 1-436/468-1021 PathoSeq: 437-467/1022-1122

ATGTCTATTACAGTTACATTTCCGAAATCTCCATCTACGAAAAAACGTGCACCG
GCATTTGGAATTGAGTTGGAGTTYAG
TCAMCAAGSCAGTAGCGATGGTGCTATAGAGAAAAGCGGCATTGGCAGTTCCT
GTGTTTAGCGTTGACAACCAAGACTWT
GTATTKATAAGAGAYCWTGCCAAGTACTGGGGCTACCCTTCATCGTATCAATT
GATTGTCAAGTTGGTCAAATGTGCTAA
CATTGAAAAGTCGCAAATCTTAAAGACCGATAAGGATTTGAATAGAGAGTTGT
TTGAGTTGGATTTGATTGAAGAAGCAG
ATACAAAGATTGATCTTTTTTATATTTCGTTACCCTTGGTCTATTCAAGAATAGA
AAATAAGAAGGTTTTTTATGTTCTG
CGTGAACCAGAACAGCCAAAGGTGTCGAAAGCMCCAACACAAGAGAAACCAG
CAAGTGTGGTTGCTGCAGAAGAAGATGA
CGATAATCTAGATGATGATGAGGAGGACGAAGTGGATGAAGACATGGATGAA
GATAATGATAATAGTGGGGAATTGTCTA
AAGGATACAAGCACATGCACAAGGACCATCCAAAGTATATAAATGACGATAG
GGTTACTATTGGACAAGTGTTTCATCAA
TACGGACTTGACCCTTCGACACCATTAACCCATTCACTTTTCAATAGTATCAAC
TCAATGTCGAAGCTAAACTATTACAA
GAATTTTGGAGTTTCAGGTTACCGATTTCTTCCCAACAGCAAGTTATCTTATGC
AGAACGAGAATTGGTGTTGAATGCCA
ACAACTACAATGATATGCACATTAACGAAAAGACAGAATCCAAGCCGAAAAA
GAGTTTCCGTAAACCCATTGGAAAGTCA
AAGAAACATAACTTGCAGATTGATCCGAACTCCATAGATTTAAGCGAGTCAGT
GATTCCGGGACAAGGGTTTATACCTGA
CTTTAGTATCCACCTATCTTTGCAAAGTCCCTAATTATTATGTGACATCAACCC
ACCAAAGTCTCCCGCTGTCGTTCAAC
ACAAAGAATCTTAATGCAACTTCGAACTCTTCGTATTTGTTTAATGATAATGTC
AAGATAAAGTCAAAAAGTATTCAGAA
GTWSGTGTTCAACAGCGATACCGATAATTACCATCACACAAAGTATTTCTACA
CCAAAACCTACCGTGGTCCAGGGTCGG
GGAATTACAAGGATGGTGCATTGATGAACAAAATCAACAAGATACATCTTTCC
AGTAATAAAAAGCCGCGCCACAAGAGA
AAGGTGTCGAACAATAACAGGTACAACAAGAGTTTAAAGGGGTTAGTCCACG
AAAAGTTTGACAAGAACTTTGTTGAGTA
CTTGCTTTCTGAGCAACGCAAGTATACCGAGGACTATTCCAATCTTGAAATTTT
ACACAATAGCTTACAGTTTAATGTTC
TTTTGAATACGTATCGTGGTGTTGCCCAAGAGACATGGAATAACTACTACAAG
TTTAAATTGATTGATTTCGAACAATTG
AAGGCTTTGCAAATGGAGGCAAATGAGCTTGAGGAGAGAAAATTGGATGCTG
CTAGACACCAACAGTGGGCGGAAGAAGA
GAAGCTTTNCCAAGAAAGATTGCGTTTAGTATTTGAAGATGAACGGACGAGTT
TGAGCAATTGCAAAGCGAGTTTGGTCA

Fig 1

GAGAAAGAAGGATTTGGAAGAGAAATTGCGTCGCCGTCAGCTANANGCATCTT
TGANTGATAGTTTTGAACTTGATAGCG
AAAATGACNATGAATCTTGACTTGNCCAAANTNAACAAGACTT

Fig 1 (cont)

>214c_cpL1 290bp inhouse:1-290

GAATCNCANACTCGNCACNGCTCCCCAAAAAGGCCAACGTTCGTGCAAAAGGC
TATACTGGTGATATCCACGCAGATGAA
GAGCAAGTTTAATCAACTCTTTGTCAATTAATGCTGTACTTGTTTTCATTTTATT
TGCTGGCATTTAAAGAATACCCATA
GTTCAGAAAATAAAATTGAAAAATTTAAAAAAAAAACGCAATATCATTCATTTT
TTTTGTTTTTTTGACAATAATATTAAT
ATGTAGTTACCAATGTTTTTAGATTTTATATGTTTTGAAAAAATAGTTTG

Fig 2

Sequences with unknown function, C. albicans sequence present in the public domain (ALCES)

>113g2 638bp in-house: 1-638

CTTATTCGGTTCTAGTGTCTCAATTGGTTATCCATTAACATCTATTCCCAACTCC
ATCATTATTGGCAATAAATAAATGG
GTGTTATATCTATTGGTAATAACTAAACTGGTGTCAATTCAATTCCAATATGGT
CATGACAATTGAAAGTGTTACTGTTC
TGGTTTACATATTCTACAGGTTACAACTATTGATTGGTTAGAAGTTTGGTTTCA
ACATCACCTGTTGCTAAGAATAAATG
TTGGTCATATCAATTGAATCATTTGTTGGTGTTATGGTAAGTAAATGCTGGTTA
TATCTATTATCTACAACCACCAAGTG
ATAAATGCTGAACCGTAGTCACCAACTGTTATGCTGGTTGTATCTATTGACTAA
AACTACCCTAGGGATAAATGCTGAAC
CGTGGTTACCAACTGTTATGCTGGTTGTATCTATTAACTGCAACCACCAAATGA
TAAATGCTGAACCATAATTACCAACT
GTTACATTGCTGGTACTACATTAAGAATAAATGCTGCATCTACAAGTACCACCT
GTTGTGTTAATAAATGCTGCACCTGC
TAGTACAACTGTTGCTGGTCATGATAGTTACTACACATTACACACCAGACAGTG
GCAAACAAGGTTATGTAGAAACCA

Fig 3

>113g4 844bp in-house 1-844
ATAGAACTGTTTGATATACAACTATCTCACTCCCAATTGTGACTTGAATAAATAATAATACCTATCACCTAGTAATCTTT
ATCTTAACGTAATCTCTGCAAAGCACAATCAATGTATAAAAGCATAAAGATAAAATCTTGGTGAGGTTTAAGTTCATAAT
TATAATGAACAACAATTACTAAAAGGGATGGTATCAACAAATTATAGGCTAGGTAGAACCATAGTGGCTGTTCGGGAGTT
CGGGTAGTTTGGGAAGGTTGGGAAGGTTGGATAGTTTGAGAAGGTTCCGTGGCTGATTCTAAATTAACAGAGAACGATAT
AATGTACAAAAAAACATTCAGAATTTTAAACAACCTTTATATATATATATTAAATGCTCTTGTCATCAACTTGCCATTGC
TGTTGATGATGCTTTCCTGTTAAATATACCTTTAAGAACCAGATTCACTATCTCAACTAATATTAACCCTTATACTTTTT
GTTTTGACATTCCATAATGACACAAAAGATTGTGAAATATTTTTAGCCTCAAGGGGATTCTACTCATTCCATCTCAAACA
CACATTCTTTGTATCACCAATACCTTTTGCTAACAGAGGAACAAAAAATTGACACGCATGTCATTTACCCTATAGCACTA
TCACTACAAATCAAAGGATTTACAATAGTGGGAATGTCAAATCATGTATATTATTAACACATTACACATATTTATTTTCA
GGTACATAATACTCAATATCTAAAACTTCAAAATGGTACTGTACCTTAAACTTTCTCCTTCATGTCTAGTTGAATATTAT
ACTTGCTAATGTCAAAAATCATGTCTTCACACATTCCAGGTTGT

Fig 4

>117c_af 623bp in-house: 1-623

AACTGTCCTGTGAAGACGAACATCACAACCACAATCATGGTCATAACCAAAAT
CACAATCATGTTGCTCCTATTCCTACA
ACAGCTGGACAATCATTAAATAATAAAATTGATACATCTAAAGTGACAGCTCT
CAACATGGCCAACTCTGCTGACGATCT
AGCAAAAGTTTTCAAAGATTCGACTAAAAAATATCAAATCAAACCAATTATCA
AATCAGACAGTGATGAACAAATGATTA
TCAACATTCCATTTCTTAATGGTAGTGTCAAATTGTATTCGATAATTCTACGTAC
CAATGGGGATTTGTATTGTCCCAAA
ACAATAAAATTATTCAAAAATGACACATCAATTGATTTTGATAATGTGGATTCG
AAGAAACCAATACAGGTGTTAACTCA
TCCTCAAGTTGGTGTTGCTAATAATGATAGCGATGATCTTCCAGAGTTTTTGGA
ATCAAATAACGATGACGATTTTGTCG
AACATTATGTGTCTCGACATAAATTCACTGGGGTAAATCAATTGACAATATTTA
TTGAAGATATTTATGATGAANGAGAA
GAAGAGTGTCATTTACATTCAATTGAATTGAGAAGGGGAATTCACTGAATTAA
ACAAAGACCC

Fig 5

>15c1 977bp in-house: 1-977 bp
TTTTTTTTACAATATAGTTAGATCTCTTTTTAAAATTTGAACACAAAAAAACAAAAAGTAAACTACTATCACCACCACCA
CCACCACCAAAACATCATAGTGGAACTTAATTGAAGAATATATTAATAACCATTAATTATAATAACATACTCAAAAGGAA
TAGGAGTAAAACCTTTATATGTAAATTAATTAAATAGCAAAAAAAAAAAAAAAAAAAAGGAAAGATTTCAACAAATCTTGTA
ATTAAATTAAATTTCATTTCATTTCTTGAAAGTGATATAGTCGTAATAGCAGTAATATTAGCAATAATATTAAATAAAAC
TTTAAAAATAACAATTATAATAATAGTAATAATAAACGAATTTAACAAAACAAAAAAAGGGGGGGGAAGACAACGAATAT
AGAAGAAGAAAAAACAACAAGACGGGTAGTAGATATATCTGGCTTAAAAAAGCATATCTAAAGTACAGCAAACACATAAT
GCAGCAAGACAACCCATTAAACAAGAATCATTACCTCCAGAACGTGGTTGTTGTTGTACATACATAGGTTGTTGTTGTTG
TTGTTGTTGATAATATCCACCACCACCACCTGGCTGTTGTTGATAATATCCACCTTGTGGAGGTGGTGGTCCATAAGCAT
TATATCCTTGTTGTTGTTGATAGTGGCCATGACCACCACCACCACCACTAAACATCCCTCGATCTTGTGTTTGTTGAGAA
TAATTGGGTTGTGATTGTGGTACATAACTTTGTTGTGGTTGTTGTGATTGGGGTTGATTATTATAATTTGGTGGTGGACC
ACTAGGTTTACCGAAATATTCGTCTTTTGACATTGTTTATATTTATAGGTGGTGTAATTTGGTGTGGTGTGGTTGTCTAA
GATTGAGTATATAGAAGTTGGAAAATTTAATAACAATTAATCTAAACTTGATATAAGATGGATTAGCAATGATAATGAAG
AAGTAAAGTTGAATGTG

Fig 6

1 QQSYVPQSQP NYSQQTCDRG MFSGGGGGHG HYQQQQGYNA YGPPPPQGGY

51 YQQQPGGGGG YYQQQQCQQP MYYQQQPRSG GNDSCIMGCI AAICVCCTLD

101 MLF

>17g1 731bp in-house:1-604 public:605-731

GCTGTAGTTTTGCTTCCAAAAGTTTGATCTCGTCATCAACATCATTAACTTCATC
TAATAAGGTGAATAATTTGGMTCCM
KGYTCCACGTGSYYGYATCACTTTTWATAGATTTCACTTYGGACAAWACTTTA
TTTCTYYGYYGATCCCATTTCYKGAMA
GATCCGTGTAATGTTTCKGCGYMAGACATGTCTTTATTATATMGTTCATTTAAA
GAATAGTGACTCTCTGACAACTGATC
AAAGGTCKGTARAATCCTACTTCGTAATTGATATATATGATTATTACCACTCTG
TAGAAACTTGCCAKATTTGACTGAAT
CTTCGTATAACTCTKGTGTGAGCRAKTTTACTCTGTTAGATAAATACTCGATTG
GTGAKKGTGAAKTGTTGTCMTTTGAC
TGGTACRKGGCTGCRGGRARAKKGATRGATTTKATCATCMAACTGTCCATGGT
ATTRKRTAACAGTKCACTTYCTTTGAT
AGAATCAATTAAAGTTGTGGTAGTCACTAGATTGGGTTTATGATTGTCGAGTAA
GTGTGATAGTTGCTCATCACTTATAT
GCTTATCCAGAAATTTATTGTACAGCACCATGTGAGTCTTTTGTAGCTGGTTTA
TATACTTTATTTAANATGAACTCTTC
GGGATCGAGTTCATCTTCATCTTCGGAGGTGGAAGCGGGGATAGAATGTAAAC
GTTTGATAGGGGTGTCTTCTTCTTGAT
AAGGTCCCAGA

Fig 7

>207g4 769bp in-house 1-759
GCAAGATCTAAACTCCAGTTTTTTTGGTGTAATGTTACACAAGCAAACAAAATATAAATCGAAAAAGCCCCAAATAATTCT
CTTCTACAAATTACGAAAAATGTTTCACATGTATGAAAAAGCTTTATCTATACTATTTCTCCTCCAACTCTAGCAGTGAG
AATGATACTGATATCTCCTATTAGGATACAGTTATCTATTATAAGTATAATAATAATCATGGAGATAAATATATATTAAA
TCGATGGAGTTAACGAGAAAAACAATACAACCCATTTTGCAGCAAAATGAGACATTTCACAGAAAAAAAAACAAGAAAAG
ACAATTACTCCATTCAAATAATTCCACAATAAAAAAATAACAAAGAACAAACGTACTAACAAAAAACATCACTAATTTCA
CTTTGAAAATCTTTACATACTCAACTTCTAAAGATTAATAATAAGCGATGCATATTCATCAGAATTTAGTGTATACAATA
TGCAGGTGATTATGAGCCAGGTGAAACAATTCTTTACTAAAAATCTAGGAGTTGTTTATATACAGTATTTTTGTCTAAAC
CTGTCTCTAACGTATACAAGATAAGATTTGTAATCGGTTAGAATAACAAGAAGGTGTGGTTGTGGACTTGGTGGTGGTGG
CAAATTTGAATGATATATTGTTTATCTCAAGTATAGCAAATACAAGGGCAAAAGGCTGCAACAAAACAAGAACTTGGATT
GTCGCAATTCTCTTCACCCTTTCAGAATGTCCTCGTGTATGTGATCAAT

Fig 8

>222g8 543bp in-house: 143-543 public: 1-142

CTACAAAATGAACAATAGACAAGTTCCTAGTAGAACACCACTAGACACCAACA
CTAATCCATGTACTTTCTCATTGGTAG
TACCACTTTCAAATATCCAATCAAACACATCAAATTTGATTGTACTGTTCTTAT
TTTTGAATGAGCTCAGNTGNNTTCTT
CTGCTACCATAAAAGAATTTGGGAATTTCAAATATTGTACTTTCAAAAGNNGAT
AATGCACTAGTAGAAGTGCCAAATAC

Fig 9

TGTTGTAGTAGGTTTCTTTGTCTTTGGTGTATTAACAGATATTACTTTATTGTCT
TCAATATTGATTTGTTCAAAAGGAT
CCTCAAAATTTGGCGTCTTCGTGCCTAATTCTGAATTGGATAATAGTATTTGAG
CCGTAATCTCATCACTATCGNCTTCT
TTACCATCATCTTCGTACCTATAATAGTAATAATAATNGTCAGACAATTCAGTT
TCACCACTATNAATTTCAGAATCCGT
TNCGTCAACAAGATTTTTTAAAATAAAATTGTCAAACATTGACGNTGCAGTAG
NGGTTGGAAA

Fig 9 (cont)

>222g9 804bp in-house: 1-575 public: 576-704 PathoSeq: 705-804

TTCCAGAGGCAACAAGCGGAAGAAGCACAACGAAAGAAGGAATTTGAACAAA
AGGCCGAATTTATCATTNGCATCATTAC
TTGAAATGCGCCGAAGAGAAATAGAGAGGCGGAAACAGCAAAAGGAAAGGG
AACAAAGACAAAAGGAGCACGAAGCAAAG
AGGGATATCAGGATACAACAACTTTCAGAGCAGGATTCACGGAGTAATCAAAC
TAAAGAAGAAGAGGAAGTGTTCAAGAA
GGCCCGGTCTACTAATTCGGGAGCAGACGAGACTGGTTTGATGTCAGATAAAG
AGTTTGATGATTCTGCATATTCACCCG
ATTATTTGTTTGAAGAGAATTTGTGGAATAAACCAAATCATCCAGATACAAATC
ATAAAACCAAAAAATATACTGAGAAT
GTGGTTGAAAATCTAGATTCTCCACCAAATGATACATCTGCGTACAATTCAAGT
TTTCATGATGAAACTAATATTCAAAA
TGAGATCCAAATACCAGAAAATGACGAGTATGTACCACAGATGAAAGCTACAT
CCAGTGTCAATAATACCACCATCCCTG
CACAAAGAAGACATGAGTCACTTTCCACTTCTGAAAACAAAAGAAGGAAATTT
GAAACAGCCGACGTTGGGGTTGATGGG
TTAGATTCTCCAGTGCGGGCACAACCAGAAATATCTGCAAAATCCAAGTCTCC
GATAATCCCTGGATGGTATACTTTTTT
ATGGNACCNAAGAGACTGGAAACTCCTGAAGGCAAATTGCTGTGCAGGGACC
AATAGGTACATTATATTCCCTCANGGGG
GNCC

Fig 10

>226c_af1 766bp in-house 1-766bp
AACGTAATTGTTATATTTTACCAAGGTAACAGGGGACCTCATTATCATTAGTTGTCAATTCAATTACTCCAGAAACAAGA
AACACAAGACTTGTTTGGTGTTGCTATTAAAAGATAATATATAATCAGGATAAAAGAATTTTTTTGGTTAAAGAAAATTA
CAGGGACGGTAAATCATTCTTCTTCCCTATAAACCAAAAATCTTATATGTCCCAAGTTAACTTATTAGAATTCCAAGATT
ATTTACTTTACAGTGAATCATTAAACATTTTAATTGAAAGCGAGTTTAGCTCAATGTCTTCAGACACAACTGCTTTTCAG
GCACCACCAACAAAAGCACCAGAAGCCTCCATGGATCTGGGTACAATTCCCAAAAGATCTCCAGCAAGATTGTTTCAAAG
GTGGATATCATCATCATCATCAAAAGATAAGCCAGTATATGCAGAAAAAGCCCTTCTCAAGAAGCAAAACATAGCACCGG
AACCAATAAAAATAACTAAACAACAAGTACCAGCTAAACAAATAGGTACATCTGAACCATCGTCGCCTCTAAGTGTGGCT
TCGAGTCATGATAATTCATGTTCCGATTCAAGTGCAGCTTCTATATTTTCTGATTCTAAAAATAACAATAGTATGCAAAT
GTTACTCACAGATGATATAGAGGACATATTAGAGGACATAGACGATGCTGAGATATACGATGCTGAGAAGGTTACCATAA
CATATATAAGTTCTAAATCATGCTAATACACATTATTAATTATTTG

Fig 11

>233c_cp1_full 500bp in-house.1-500 bp
GAAAAATCAAACAACAACAACAACAGTAAGCCAAGTGATAGTACCAAATCTACTTTAGCAAATGATGAAACAAGAAAAAC
ACTTGATCCTAAAGGCGTTGGAAGCACTACAACAGGTGATAAAGACACAGTTTCATCAGACAAAGCATCTCTGCCAATTG
AAGATAAAGAAAGTTCACCATCCCTAGCTGGAAGTTCAACATCAACACCAAGTGGAACTGATAAAAAAAACATCTCCTAAA
AAATTAGTTACCAATGCTGTCAATAAAGTTGAAAATAATGATGATTTCAAAAAATTCATTAATGAGGCTGAAAAGGAAGC
TAAAAAATCCAAATCTGGATTGAAAAAATTATTTAACAAGAAGTAGAAGTTGTTTAAATTGTTTCGATATAAATTGTATG
AATTCCAGTTTTTTTTATTTTTATTTTATTTTATTTTAGTTTAGTTTTGATTTCATTTTTGTTTTTGTTTAATTTGCAATA
CAATATAATGTTTATTTTTT

Fig 12

>22g3 (5') 535bp in-house: 1-535

AGGTTCCAGTTACCAATTTAGGAAGTGTGTTGCAAGCAGGGCTACCAAATATG
GGTGGCAACACATATGGTAGTAAGTGC
TACCAATGTGGGTGCAAAAAATTTTGCCAAGTAATTTGTATGGCAATAACAGA
AGTGTTGGCGGATTCNAACTGAGGAAT
CTTTGGTGTGTAAAAAAAAAAGCAATAGCGACTACGCTACAANAGGCAATCNAT
TATTATTATAAAGTGGAAGTTATATAT
ATNTTCTCGGGGGGGGGGGGGGNTTNGGNNTCCCCCCCCCCCCCCCCCANNTTT
TNTCGGCCCNCCCACCNTNCGGCCTTC
TGGCTCCCCCCCCNNCGGGCCNCNNGTAAATNCCTCCACCCNGGGANAANGGNA
AANGGGGAACNANNAAGGGGGGACNNN
NCACCCNATGGGAGGGAAAATCCCNAANNTTTNCCCCCCCNNCCCNGCCNAAN
CCNCNTGGGGNGGGCCAAANNCNGGGG
GCTNCNCNCCCTNCCCCCCCGCCNTNNCCCNNNNTNCCNNCGANCTCTNNGNG
GC

Fig 13

>22g3 (3') 426bp in-house: 1-426

CCCCCATATAACGTTGTCAATAGCAATACTCTGTCGCACCCATAGTGTGCACTT
CTCGGTGGTATAAAAAAAATTTTTTC
TCCCAAAAAAAATCTTCTCCTTTCCACCACTTTTTTCTTCTTCTTCCTTCCCCATT
CCCTCCCAAATCCCTCATTTTCCC
CATTTCCCCTACCCTCCTGGCCCTGTATTCCAAAATTTTCTCGGGGNTACGCCC
CGAAGANAACCTCCCTCCCACCCACC
CATCTTTGTCNGGNTTCGACCTTCGGCCTCANGGCTCCACCGTCGGGGNTCTTG
TATATTTGTAGACTCCNGGAAAAAGG
GAAAAGGGGAGGAAGAAGGGGGGAAAAAAAAAAAAANGGAGGGNGAATCCTT
TTTNTTTTNCCCCCCNTCTCAAACCNAAA
CCCCNTNTGGGNGGTCNAATTAGGGG

Fig 14

>24gG 522bp in-house: 1-522

GAGCAAGCNTGGCCAAAANNCCCACCTTGTTTTGAATAGGTTTGCGTTGTAYA
GGCAGTYAAATGTGTTTTTTKGGCTTG
ATTTTGAGAAAAAGTTGACTGAAAAAACATGCAAGAAACGGGGTGATCATGA
AAATAGTCACACACAAAWWGTCAAAAGA
CNATCACMGAMGACTCAGAATGAGCAGAAGGACTTGTCTGAATTGAGTTTTCA
ATTGTTATTTAGAGTTTTAAATTAGAG
TTGTAAATTTTTGGTCAGATTTTACGAAAAAACTGGCTGAAAAAAAAAACGAGT
CAGGGTGATTGCMAAAGAACAGAAACA
ATMSATAATCTTAAATTAAGGTAGTAAAGGCTCTGTGAAGTAATTTAGAGTTTA
AACASGGGGGGGCGAGTCAKKKTTAG
AGTTGTGAGTCTMAAGTTTATTTGGCTAGTGAATTGACTGGCMMGATTGTTAA
ACGTKGGGTAGRAAAAGACACCCTCMC
SCGTYRMCTNTTGGCCAATTCAACNCGTCCCNGGANACCGCC

Fig 15

>28gK 475bp in-house: 1-475

CCCCGTTAACCACTTCTAGGTATACCATTTCATCTGACTGAATAACTGGTTAGT
CGATTTGTTGTTGAAGAAAAGTGACC
ACCTAGTTTTTTCTGCCAACATTTTTTGCGATGAGCCGTCGACGCGTTGTCTTTT
TCTACCCCACGTTTAACAATCTTGC
CAGTCAATTCCCTAGCCAAATAAACTTTAGACTCACAACTCTAACACTGACTCG
TGCCCCCCTGTTTANNCTCTAAATTA
CTTCACAGAGCCTTTACTACCTTAATTTAAGATTATCTATTGTGTCTGTTTGTTT
GCAANCACCCTGACTCGNNNGGNNT
TTCAGCCAGTTTTTTCGTTNAATCTGACCAAAAATTTANGNCNCCGATTTAAAA
CTCTAAATAACNATTTAANGNCGGNT
CAAANAGTCCTTCTGCGCGGNNGANTCGTCNCTATTGTCTTGNGANATGATGT
GTGTGNCTATNTTCAAGAACAT

Fig 16

>328c1 681bp in-house: 1-681

AACCTAAATATGCCCGATTTAAACAAGTTGATGTATTCACCAATGTCAAATATT
TGGGTAATCCAGTTGCCGTTATTTAT
GATAGTGATAATTTAACCACTCAAGAAATGCAAAAAATTGCTCGATGGACAAA
TTTATCAGAAACAACATTTATATTGAC
TCCAAAATCATCAATTGCTGATTATAGTATTAGAATTTTCACTTCTGGTGGGAA
TGAATTACCATTTGCTGGTCATCCTA
CTTTAGGTACTGCATTTGCATTATTGGAAGATGGTAAAATAAAACCAAATGAC
AATGGACAAATAATTCAAGAATGTGGT
GCTGGATTAGTGAAAATATCCGTTGAAAAAACACCTAATAATAATAGTAATGA
GTTGCCGTTTTTGTTATCTTTTGAATT
ACCATATTTCAAATTTCATGAAATTGATGACAAAGTAATCGAGGAATTACAAC
ATTCATGGAATGGAACCAATATTATTG
GTAAACCCGGTACTTATTGATGCTGGTCCAAAATGGGCAGTTTTCCAACTTGGC
TCCGGTAAAGAAGTATTAGACTTGAA

Fig 17

TGGTGATTTAGCACAAAATGAGAGATTAAGTTTAGAAAATGGNTGGACAGGNA
TTGGGNCTTTGGNAACATTATGAAAAT
GGGGATCGGTCCAATTGAGAAATATTGCTCCTGCTGGTGGA

Fig 17(cont)

>33gK_part1 1171bp in-house: 1-588 public: 805-1171 PathoSeq: 589-804

TCTAAAATCTTCAGTNCCATNCAGAGTTTTAGGGAATGTTACAGACTCAACTCC
TTTTGCCATGGGGACATTAGGTTCAA
CATTTTATGCTGTCACTTCTGTTGKCWGATCTTTCCAAATYTAKGASKKGGYW
ACATTACATTTATTGTTTGTTTCCCAA
ACTCAAACTCCTTCAAGGAATWACATGTTTGGCTGCACACCATCAMTTTGTCT
ATGCATCTKATGGTGATCGTATTGGTA
TTTTTAGACGTGGTAGATTAGAGCATGAATTGGTTTGTGAAGGGAACTCTACAG
TYAACCAATTATTAGTATTTGGAGAA
TACCTTATTGCTACCACATTAGAAGGTGATATTTTCGTATTTAGAAAAACTGAA
GGAAAGAAATTCCCMMCTGAATTATA
CMCTACMATCAGAATAATTAWTYCTTTAGTTGAAGGAGAAATTGTGGGATTA
ATTCMTCCACCTACGTATTTAAWWARWR
TWRTTGYTSCWMCYACYSWWTCTGTGTTTGTTATAAATGTGAGAACTGGCAA
ATTATTATACAAATCCCGGGAATTACAA
TTCGAAGGCGAAAAGATTTCATCAATCGAAGCTGCTCCAGTTTTGGATGTAATT
GCTGTTGGTACATCTAATGGAAATGT
ATTTTTATTCAACATTAAAAAGRGGGAAAGTGTTGGSCCAAAAAATTATTACTT
CTGGAACTGAATCTTCCTTCGAAAGT
TGCCTCGATCTCYTTTAGAACAGATGGAGCACCTCATTTGGTTGCTGGTTTGAA
TAACGGAGACTTATATTTCTACGATT
TAGACAAGAAATCACGTGTTCATGCTTTGAGAAATGCCCATAAAGAGATTCAT
GGGGGTGTTGCAAACGCCAGATTTTTG
AATGGTCAACCAATAGTATTATCAAATGGTGGTGATAATCATTTGAAAGAATTT
GTTTTTGATCATAATTTAACCAGTTC
GAATTCATCCATTGTTCCTCCTCCAAGACATCTCAGATATAGAGGTGGGCATTC
AGCACCACCAGTAGGTATAGAATTTC
GTCAAGAAGATAAAACCCATTTTTTATTGAGTGCTTGTAGAGATAAAACATTTT
GGACATTCTACTTTGAGAAAAGATGC
TCAAGCACAGGAATTGTGTCAAAGATTGCAAAAATCTAAGGATGTATAAAG

Fig 18

>33gK_part2 1001bp PathoSeq: 1-1001

ATATCATAACCGCCCACAAGGATGAAACTTTTGCGAGAACATGGGATTCAAGA
AATAAAAGAGTCGGTAGACATTTGTTA
AACACTATTGATGGTGGCATTGTGAAATCTGTATGTGTGTCTCAGTGTGGTAAT
TTTGGTTTAGTGGGATCATCACTGGG
TGGTATTGGATCATACAACCTTCAAAGTGGATTGTTGCGTAAAAAATATGTTTT
ACATAAACAAGCTGTCACCGGTTTAG
CAATTGATGGAATGAATAGAAAAATGGTTAGTTGTGGTTTAGATGGAATTGTG
GGATTCTATGATTTTGGAAAGTCTGTC
TATTTAGGCAAATTACAACTTGAAGCACCTATAACATCCATGATATATCACAA
ACTGTCTGATCTTGTTGCTTGTGCCTT

Fig 18(cont)

GGATGATTTGTCCATAGTTGTTATTGACGTGACTACTCAAAAAGTCATAAGAAT
ATTATATGGTCATACCAACAGAATTT
CAGGAATGGATTTCTCGCCTGATGGGAGATGGATAGTTTCAGTTGCATTGGACT
CCACTTTGCGAACTTGGGACTTGCCA
ACTGGTGGTTGTATTGATGGGGTGATTTTACCAATTGTGGCAACTGCAGTTAAA
TTTTCTCCTATTGGTGATATCTTAGC
GACAACACATGTCTCTGGAAATGGTGTATCCTTATGGACTAATCGTGCCCAGTT
CAAGCCTGTGTCCACCAGACACGTAG
AAGAAGATGAGTTTTCAACTATTTTATTACCAAATGCTTCTGGAGATGGCGGTT
CAACAATGCTAGACGGGTTTTTGGAC
GAGGATTCTAATGAAGACGGCACTATTGATGAACAGTATACATCTGCTGCTCA
AATTGATGCATCCTTGATTACTTTATC
ATCAGAGCCAAGATCAAAATTCAACACTTTATTGCATTTGGATACCATTAAAC
AACAAAGCAAACCGAAAGAAGCACCTA
AAAAACCAGAAAATGCACCTTTCTTTTTACAATTGACTGGA

Fig 18 (cont)

>33gK_part3 414bp PathoSeq: 1-414

AAATTGCGTAAATTGGATACAAACGGTAACCACGCATTTGAAAGTGAATTCAC
AAAACTATTAAGGGAAGCTGGAGAGAG
TGGACAATTTGAAAGATTTTTGACTTACTTACTTAACTTATCTCCTGCTGTATTG
GACTTGGAAATTAGATCACTTAATT
CATTTGTTCCATTGACTGAAATGACAAATTTTATTCAAGCTTTAAATGCTGGTTT
GAAATCAAACGCAAATTATGAAATA
TGGGAAACTTTATATGCCATGTTTTTCAACATACATGGTGATGTTATCCATCAG
TTTGAAAATGAAACTAGTCTTCATGA
AGCTTTGGAAGAATACAGACAGTTAAATGATGAAAAGAATAACAAAATGGATT
CTTTAGTGAAATATTGTGCTAGTATCG
TAAGTTTTATTAGT

Fig 18 (cont)

>35gK 1334bp in-house: 146-669 public: 1-145 PathoSeq: 670-1334

ACAACGTATAATCGACAGTTTACTATATCTGCTGACTTCAAAACCAATGCATTC
TTCAAGCGTGCTCTGTCGATTTCTAT
CATAACATCCACTTTCCGGNGTAATCGGATTACTAAAGCCACAGAATCAAGGT
GAACATCAAGCTTCAACTTCTTTCTTG
GTCCACGAATAATTTTAATTTGGTTMTTSKKGSMAMKGCTTTCTACRGTAGGTT
TGAATCTTTCCAACATTGTCTTTGCA
TAGAAACMGCACCAGACAAGAAACATGTCCACTCGACCATCAACYTSKGGGT
AWWGACAAAGTWAATCTGTCTGGATCCT
TTTCATCCAGTTTCCCTGCATKGGAWACAAGTNTGTCCCGCACAGTTAAGACT
GTTTTTATTTTSKTGGTATTAGACTCA
TCAAGTTCCGAAGGAGAGGCATCATTTARGGGWATAGACTCCGCTGAGTTAAT
ACTGGATAAATCACTTATTTCAGATTC
ACTGACTTGTWCTTCAGTGACCTTATCAAAATCCTCAATGTACTCSGARGCGTW
TTCMCTCMATGTGAAGGCTTTTAAAA
GGGCAACRCTGGTTYCAAAATGCTTTCTTGCRAGTTTGTACKTGACAGAAAAA
TCAAAAACYTTGAAAGATATACCTCTT

Fig 19

CTAAAGTCTTTTAAATCAATTTCTTNTCCTAATTTTTCATCATATAGCTTATGAC
TTGGCAAACCCTCCTTACATACCAT
ATCCATTACAATGCTAGAAATGTCAATCTTCACTGACGATATAAAGGATGGAA
GAACTTCAAATAATTTTATAAACTCAG
GATTGGCTGGTGTATCTGCTGCAGGAGCTCCAGATTTATTGTCCATTTGCTCAC
TCCATGGACATACATTATTAACGTCC
ATCTTTTTCCATTCTTCAAATTTCTTCGGTGAAATAAATTCGTTGACGRWTTTTA
AACAGACGTACAATGTGAAAGATAA
GATCATTAGCAGAGAGCAATTCGAGACTCTTGCTTGAAAGTTTGATTGACACG
TTTTGTTGTAACATATTGTAGGTGGCT
AAAAGATTGACTTWRGTAAAATGRAACTTATTAACCCTGGGCCCTCACATTTC
ACATTTTTCATCTTAAACAAAGKGGTT
CAAAGKGGAACTTGGTTTGGATCCYTTAWTGGAAWATTTCYCAGKRAATACTT
TCAAAATCAACTCCAGGAGAGCCACAG
TGATAATTGAATTGGATTTAGATAAGCGGTTAAACTTCCCAATTTCAGTTTTAC
CAAACTCTGGTAAATGAAGGTTAAGT
TTTGTGTCCACCACAACAAGTTTACTAAAAACAGCCTTGAGCATTTTGGAGGCA

Fig 19 (cont)

>36g2 (5') 520bp in-house: 1-520

CGTATAGAGAATAATCCGTTGAAATTGATTGTTCAATCATTATTGTATCTTTTCC
CTTTTTTTTGTTCTAACCATAATGT
TAGAATAATTAGAAATTGTCTAAATATATATTCAGTTTAACAAAAAAACAGAAT
GCTTGCAATAAGATTTGATTTCTAATT
ACTAATCGTTAATATTTAGTTTGGTGGGGTTTTATTTATCGAAGATGTAGCATT
ATTTGTATCNAATAGATAAAGAAACT
TGAATTAAATGGCNTAATTTGTTGCAATAGTAAAAAAGAAGAAAAGTGGTAAG
GAGTGAGTGAAAATATTTTTTGCCCCA
ATTTGAGTNGAAATCTTACACCNAAAAGTTTGGACNAAAAAGTTTTTACTAAA
ATCTGANAATCTNCCTGAATAGAACCG
ATCATCCNCATNTCCGATTTCNTGAGGANAGATAGTGGCCCCACCTCNTGGTG
ATTAGAAGGAGCNCCCATGTTTTACAA
TATCTATATCCAGAATAACNTGTTTGTGACCTCNCCCCNG

Fig 20(a)

>36g2 (3') 472bp in-house: 1-472

CTCTATATATAGTGAAATATAACATCAAATAATGTACAAAAAAGTATAATAAA
TTGATTTAGAAATGAGAAAAAGAAAAA
AACTTGAAGTAGTGAAGATATATTTGTTGGCTATCTTTCTTGGTATGGCTCAAT
TCAGCCAATCTTGGATGAAAGGTTGG
AGTTTTAGTTTCGTGGTTTATTGATTTGTAAGTACTTTCGGGCTAGAAAGTTNA
CAAACATGATTAATCTTGATATANAT
ATTTGTTAAACATTTGGTGCTCCNTCTTAATCNCCCAAAAAGTTTGGGNCACTA
TCTTTCCNCCNGAAATCTGTATATGT
TGANTGANCCGNTCCATTCCTGTTNANTTTCNGANTTTAGTTAAAACCTTTTTG
TCCCAACCTTTTGGGGTTAGANTTCN
NCCCCANTGTTGCCNNAAATATTNCNCNCCNCCCTNCCCCTTTCCCCNTTTTAC
NAATGCACCAAGTAAGCG

Fig 20(b)

>38g1 1348bp in-house: 183-940 PathoSeq: 1-182 / 941-1348

TCTCTGGTATAACTTGCACTACCTCATCGCTACCCCGGATTTTTTTTTGGTATGA
TCTACACGTCCTCATCGCTACCCCA
GATTTTTTTTCTGGTGCGCCGGACACGCCCTCCGGTCCGCACCGAAAACCGGGG
TAATCTCCGTCGGAGATACACATCCG
CGGACACAAAATCAGATGAGCTACCACCGAAAATTCCGAAATTTCAAAAACTC
AAAATCCCTAAAAACAAACTATCCAGA
NATTATTGCCATGCCCTGAGGATGAGTTTAGTTTTTTAATTTTTGAAAAATGTC
CAAAACTGGTTGTGCTGTATAGGANG
GGTAAGAATTTGCCATTCTGCCCCTTTGGGTGGGTCAGTCNAAAAAAGANGTA
TCACTCTGGTTCNAACGGGAAACAACN
NAAAATGGGATTAAAMTWATCTCCAGAMCAAACTTAGCTTMWWACACCCAY
TTTAGTTGTACTSGYGWRCCMAAMMCMAA
TTTTCCATTTTGTTTGGGGANGGGAATTTARACCAAAWTTTTTTTTTTGAAATTT
CGCTMAGTGTYMAGAMCCSCAAAAG
TCACCTTTTTTCGTTTTCMMCYACGGCARARGCYCACCGGTTTTKYKTGGKGS
MCRGCCMAATTGAWTTTGTGGGTGSGC
ACGKGGAAAAACAGTTKGTTAGTGGACACGTTTTTGCAGTGTGAAACTGCGCT
CGGAGGTACTATATGCGAAAGCAGAAA
AGACAATTGCAAGAATACAGAGAGTTCTTCTCTGGGCTANNGCAATGTGTTTA
AGGCCAAGTCGACGAGTGGGGAGAGTC
TGGAAGTGATATACACATCACGACCTACTTTATACGCTACGTTCGGCATGGGC
GAGCCACTGTACGGTGGCAAGCCTGAA
CAGTCCCACACCAGATATCTAACGATTCTGTGTATGGGCACTGATGGGATTTAG
TGGATTACTAGCTGATAGCAAGTATT
GAAAACTAAAACCCGACTCGGGGGTATGCCTTGGCAAGTAGCCGGAGTAAAAT
CTGTGACTTTGCTGAGTGTAACTCCCT
CCATGGTTGGCGATGTTCGACGTGCGCGGCAGTTCTTGTCGTATCACAGTCGCA
CGGACACCACACCGGGAGAATCTTAA
GAGGGCTATATGGATGTGGAACGGTTTGCTTGCTGTGGTAAAACACTGGCGGG
CGAGCCGACGTTCCACGGACACAGCAA
TGTGTTTGCAACCAAATAAATAACTTGTACGGTTTGAACGTGTTTTTGGCTGCT
CCTTCCAGTTCTTGGCGGGAGAAGCT
TGGGCGCGGGAAGACCACTACTACGTAGTTATCTGGTTGATCCTGCCAGTAGT
CATATGCTTGTCTCA

Fig 21

>3gG 842bp in-house: 171-842 public: 1-170

ACCANTAAGGGANGAGGTGGACAAATTGAACCAGAATTGGGAGCCGCCAAAT
TGGACGGAAACCCCGGGGAGGCGGCCAA
TTTGGGCGAGAACCCACGGATGGACAAGCGGGGCAAGTGCCAAGGAAGTTGC
GTGCCAGTGAAGTGGAATGTTGGAGAGC
TGGCAATGAACTGCCCGGCAAGTGATACCAGGAGATAGGTGTGTATAGATTAT
NATGGAACGCCNATTTTTGCAGTATCA
CGCGTAATAAGGACAGCAGTTGGACATCGGTACATGAGAGAGCAATGTNAGTC
TTGATANTAATGAGCCGTGTTGAAGTA
GTATTTTAATCNAATTTTACTCCCAAAAGGACAATGGAGATCTGGAGATAACN
CCACACTAATCGGTTCTAGACATAGAC

Fig 22

TAANCCTGAAAGGGGGTACTACAGCTTGTTTTGAAAAGGTTTGCGTTGTATAG
GCCAGTAAAATGNNNCTTTTNTTNGGG
TAGAATTTGAGAAAAAGTTGACTGAAAAAANCGCAAGANACGGGGTGNTCAT
GAAAATAGACACACACAAAACCGTCAAA
AAACAATGGAAAAGCTTCNNCATACGCAGTAGGAGGTGTCTGAATTGAGTTTG
TATTGTTATTTAGAGTTTTAAATTAGA
GTTGTAAATTTTNGGGTAGAATTTACGAAAAAGTCGAACAAAAAAACGACAAG
TCAGGGTGATTGCAAAAAAAACAGAAAC
AATAGATAATCTTAAATTAAGGTAGTAGAGGCTCTGTGAAGTAATTTAGAGTTT
AAACAGGGGGGCACGAGTCAGTGTTA
GAGTTGTGAAGTTTATTTGGCTNNTGAATTGACTGGCANGAT

Fig 22 (cont)

>480c 731bp in-house: 1-731

TTTGAANTCTTCNCCNCCTGNNNCTNCTCAAAGCGCTCCGCNTTGTGNCTNAAN
GGGCTGGCTCAACGTACTATCAAGNG
TAACATAAAACTANNACTTGGGNCATTCACGATAAAGAAACGGNGCCTGGAAT
TCCAGCAATTGNCGCTGGACTTGAGTC
GNAATCTAGGNATTGNTAGNCGAGATTTTNCATATGAANNACCTGGGAAACGG
ATCAACTGGCTTAACANGACCAGTATT
GATGNGGAGAGAAAAGTGGGACTTGCAGANTTTCTCAATANCCTCATTCAAGA
CTCAACACTTCAGAATGAACGAGAAGT
GTTGTCGNTTTTGCAATTGCCGNCTAATTTTAGATTCACCAAGGATATGTTACA
GAATAATCGAGCNNACTTGGATTNTG
NGCAAAATAACTGGTACGATGTATATCGTAAGTTGAAACTGGATATACTCAAC
GAATNGTCTAGCAGCATTAGTGNACAG
ATACATATTCGTGATCGCATTAGNCGGGTCTACCAACCACGGATTCTCGACTAG
GCAGGNCTATTGGTACAGATAAAGAA
GANGCCTAAAGAAGAAGCAGNTGGGTTCCCAATTCTTTNAGAGNNTAGAAAAT
TTGGTAGTACAGGAAGTTTCCCGATCA
AAGAGGGTGTTGGACCCAACAGNTTAANNTTTCGGCNGNGNCNTTNTCATTNA
ACAATAAGNACTTCTTNAACACCAAAT
CCNNGTTTTTA

Fig 23

>55g3 1063bp in-house: 533-1014 public: 88-532 PathoSeq: 1-87/1015-1063

TTCNCCCCCATTCCAAGATTCCCCTTGTAAGTAAATTGGTTGGAGAACCNCGTT
GGTTTAATTCCCCCGNNCGGGAAANN
ATTNGTNGTAAGACAATTCTTTCAACAATTTATGATGTTGCTGCATTTTCGTAA
ATCTGCTAGTAAGAGCACCCTAATCT
TCGAAATCAATAGAAAACTCAATCTTTTCGATTTTCTTGTGGATCACTTTACTC
ATAGGCTCGCTATAAACAAACAAGCT
CACTATGATAACATATAAGGTAAAGTATTGACTTGATGTAGATGTAATCACCA
CAGAAGGCATATCAATACCAAGTCGAA
GAGGAGCCTGTTTCGAATTATCATCATTGTAATTCGCTTGCTCTTCTTGATCCA
GATCTTGTGCATTTGGAGAAAGCTTG
CTTGACAAAATTTCACTTTCATAACACATTGTCATAACAGAAAGCTTTTCAATC
AATAAGTTGTTGGCTCCAGCCCATTT

Fig 24

ACCATTTTGGGTTATTTCTGTTCCTAGCCATGGTGGCCAATTAGATTTAGCTCC
ATATGGATTACTAATACTTAACATAT
CTGGACACCCTACAATATCCTCTTTGTTTAATACAAAAACATTGGCATCTTTTA
GTAAAATACCATACCTAGTTTCTAAC
AAGATTTGGTTTGCATTATTCCTGGAATCCCTAATGGACAAAATTTTACCTTTA
ATAGATGGTGTAGAGATGAGCACNAC
AGAATCAGGATAATCCTCNCTTTGTTATTGAATTTGAGGTGGGANCACNTCNAT
CAAATAATCCTCAACCACTTTTTCGT
CNGGTCTGGTCTCNCNTAATATTGTGTTGAAATTGTCCAGTTTCCGTTGTGAAG
TGANATCNTTGCCGGTAGANGTCTGT
GATTTAACCCCCTCCCGTGTGNAACGACCGAATANTTCTGACGTCCANNAAAA
ANCGTCCCTATCGTTGTTTGCNTTTAC
NCCATCCNCCNTTTTTTCCAGTGTTTTCCATGGACTTGTGTGNCGAGTTATTTCG
AAGCTGTGATTTCAATTTCACAAAA
TGTATGTATTTCAATGTCAAATT

Fig 24 (cont)

>58gA 724bp in-house: 281-582 public: 209-280/583-724 PathoSeq: 1-208

GTGGTGTTTTGGAAAGTAAGGTGTGATTTGCTTAACAAAGGAAAAGAACGAGA
CGAGAACCAATTCTCAATATATAGGTC
TTTCCAGGTAGGAAAACGACCAACTGTGGAAGAATGGCACTACCATTGGTTAC
CAATACAACTACTGGTTGTCGTTTCAT
CTGATAACATACACAACTACCATGTAGAAGTCATCAATTGTAAACTCAGCTAT
ACTTTATCAATATCAGCGAAATTTTAT
TTAGTTTTCGTTNGAAAATACAGTGAAAAATAAAAATCTACCCGTCNNCNTGA
ATTGTNTNCCTCTGCANCNACAAATNG
TNNTTATATTGTGATTCATTTCNNAGGCTTGATTNCCANCTATTTNTAAACACT
ACCTTTCATTTNCTACNTTCNGGGAA
AATAACNCTTGTTGCTGTTGAAAGACCAATNCCNTTGTGAGTACAGAGGAATA
CTNCCANTATNCGGCTTATANTTANCT
AAAATTACAATACATAACAGGGAACCAGACNTGTTTNCGTCNTTGATAATGAA
CAGTTNTGGTNCTNNTGAAAAGTAATC
CCNAATTTGAATGGNTCGAAAGCAACACAATAAGAGTCTTTGCTTGATATTTG
CTTCTCCAGAATATAAAATAACGCGTT
AAAAAGACGTGTTTTCTCTTCAATCGCCATCACAAATAAATTCAACAGAGTAG
TAGATCCTGTTTTTTTTCTTCGCCACA
TCAC

Fig 25

>60gK 990bp in-house: 445-752 public: 1-140/753-990 PathoSeq: 141-444

ATTACCGATCCGTCGGATTTTAAAACCACAAAATTGCCTGCATTAGCAGAGCT
AGATATTTTCATAGGGTGCTATATATG
CAAAGATCTATTGAATGCACCCGTGAGGACACAATGTGATCACACGTACTGTT
CACAATGTATACGAGAATTTTTACTTC
GAGATAATAGATGTCCGCTTTGTAAAACAGAGGTTTTTGAAAGTGGTCTAAAA
CGTGATCCATTGTTAGAAGAGATCGTC
ATTAGTTATGCCTCCCTTAGGCCTCATTGATTACGATTATTGGAGATTGAAAAG
GTGGAATCGAAGCAAGAGGTAGATCG

Fig 26

TGAGAAATCAGCCAATGAGTCAGCGCTGAATGGTAATAGAAATGTAAACAAC
GATGTTGACGAAACTGTGCGCGTTAAAG
ATCAACTGAATGCAGATAAACTAGGTGAAGAAAAAGGGCAAGCTCAACATGG
GGAACAAGTNAAACGAGCAGACTACTGA
AGTTATTCTGTTGCTATCTGATGATGAAGAGAATGGTTCTGATAGCCTAGTAAA
ATGTCCTATTTGTTTTGAGAGAATGG
AATTAGATGTACTACAGGGAAAGCNTATTGACGACTGTCTAAGTGGAAAGAGC
ACGAAGAGGACGCCTACAGACATTTTA
TCCCCAAAAGCCCAACGACCGAAGCAAATCACCTCCTTTTTCCAACCAACAAT
AGATACCANAACNCCTTCCCCCACCTA
CCAGTTNNGGCGTCNACAACTCCCACAGCAACTCCGACAACTACATTGTTGAA
AGCAAACGTCTCATCTCCATCCCAAGT
GGCGCAAAGTACAGTAAACAAGGGCAAGCCATTACCTAAACTCGATATCAGCA
GCTTGAGTACTAAAAAAATAAAAGCCA
AGTTGAGTGATATGAAACTACCAACAACAGGTAGTAGGAATGAAATGGAAGC
CAGATACTAGCATTACTATGTGATTTAT
AATGCCAACCTTGACACCAATCATCCTGTA

Fig 26 (cont)

>61gB 602bp in-house: 1-602

ACCTACNNTCACNCNNGNCNCNGCAACACCANCNTNCCNCCNAAANAAAGTC
TTCTTTGAATNAGACNTTTCATCTATTG
GAAGACTTGATCTACCTGAAGGTCTGTTTAAATTAGGATTGAAATGATGGTGTT
GTTGTTGTTGTGGGTGGGGATTGTGG
TGATCATGGTTTTGATGCAAGGGTGATCCAGGACCGCGCTGGTTTCTTTCNAAT
TCATCTAAAAGAGCCATATGTCTGGG
TNGCCGTNGTNGTTGTTTCTGAGCAATGAATTCTTGTTGGATTTGCNTGNTGTC
ACGTTCATNTTCTTGATTGAATTTAA
TACCATTTTGAAAATCATTAAAATCTGCCACTNCNACAGTTACTCTTCCTTTTGT
CTTANGATACAAATNNTACCCTTTT
NCACGTTTGGCGTTTGTGNCCAATGCAAACNCNGTNCCCCGGGGNGNGGGGNC
CCCCCCNGCCNTTGTCCANANCCTGNT
GGNTAGTTATNTTCGCGCTTACNATCCCCCCCACCCNNGCGNGNNCCGCCCCA
CCCCCAACGTNCNCCTCCTCTCCCCNC
NNCTNGCNGCGCCTNTCNGGGCACCCCTTCNCCNCNCCTCNC

Fig 27

>62gB 539bp in-house: 101-539 public: 1-100

ATAATAGAATCTGATTTGAAATAATAGGAACCAACAAGAACAAAAAAAAACAA
GAAAAAAAAAAGATTTGTATAAACAAAA
ATAGAATCAGAAACAAAAAGCTKTAGTTTGKGAAGAAATTGAAACAATCGGA
AAACAACAATATCAAACTGMTGCCCAAT
AACACTGGTATGTACCTAGATGGATTACCAAGATCTACTACATAAAATAATAR
RGGAGTTCCACTCACTCAAAGAGTTCA
AACCATGGGATAGCAGTGTTTTGTATGAGACGTTACTACGATCAGTATTAACTA
CTTTGATCGAACTTTTGGGCMTAGAC
AATCCACCCAGTTWTCTWCMCCTCACCACCMACMATGATAGTTATAGGTGAA
TTTGAAAATWAAATACTATGGRAATGCA

Fig 28

TTAAGCAAGTCAATCAAYSSWCMKRGCATRKTGCAATWTGCYWGWAYCAWA
GSATGYAATCSATATTACMRRCYKTKSYY
GRGRYYRKKRATACGCGAWMATWTWKAATCMMMGAGTCYWATYCTGCTGK
WWTCMAAGA

Fig 28 (cont)

>64gB 627bp in-house: 1-627

TNCANCCTNCCATNCNCCCAGGCNNNGCCACCCCNGCCGNNCCCCCNTNTTTC
CCCCCCTCCTTNGTNGCCCTCNNGGTG
GTGTTTGTGGTGTGACNAATAAANATGGTNTATCATTAGAANAGGACATTGCN
NCGGAAATGACTGTCGACAATAAAGAA
GCAAATATATACAATGGATTATGAANGTGCTAGGATGGATTTGAAAGTTTATC
TGGGTTTATTCCAATGTAAAAATTATT
TGTAATTGATATGGCTAATTATTTTGCTCNATATNTATCACAAAAAAATGATTA
AGTTCGAAATGAAATTGGCNTCCATA
TATAAAATTTCTGACAGGAAGAGAAAATTCANGACNTGTTGCCCNAAAAAAAA
AACTTTACCCCNCNTCNANTCNTGTNN
GACTTAACCCCCAAAAANAANANNGCTGGCGGCGGNAAAAAAATAGGAGGGG
GCCGGNNGTTTTTTAAAATTTNANNCTT
GAATATGAACCCAANNTTTGNNTTCNTTTTTNCCACNCCCCCTTCAAATTTNAT
TCCATGTTCCCAAGANNAGGGNGGNG
GGGGNGGTTCCNNCTTTTAAACCNCCCCCCCCCGGGTGGNGGGGNCCGTNTTNT
TTCCGGNGGGGCNT

Fig 29

>65g 441bp in-house: 1-441

TTNCTTATGTAGATGTTGTTCATGAATTTGTATGAACGGACTATGGCTAGGATT
TGGCCAATCTCGGTATTACTACNTTT
TCAAGTTCAAAGATTGGGAAACTCGTGTATTTTCGTACTGTCTACATTTTCTTA
AATTTGATAAACGCATAGTAAGTCTT
TGCTTGATATACTATGAGATGATTAGAATTAAAAAGTAGACGACTAGTTTCACT
AGATTTATTGAAGTGTCAAAATATAT
TCAGATTGGTTGCAACTGATGGTCTCGAAAATGCNACAGGATTTTTTTCCCCCA
TTTTTTGCCAATTTTTGTCCNATAGA
GTAGAAAGTACCNGTATNCNAATTGTCCCAAAAAGCGATTATAATCCGTACCA
ATATTTCCAATTTTCNTTTAAACCCTG
TTCNCCTCNGTGTTGGTTTGTTTGAAAACNTAACCANGGTG

Fig 30

>8c_cp 890bp in-house: 287-890 public: 1-124/154-286 PathoSeq: 125-153

ATGCAATTCTCATCCGGTGTCGTCTTATCCGCTGTTGCTGGGTCCGCTTTGGCTG
CTTACTCCAACTCCACTGTTACTGG
CATTCAAACCACTGTGTCACCATCACTTCATGTGAAGAAAACAAATGTCACGG
AAACTGGAAGGTTACCACTGGTGTTAC
CACCGTCACTGAAGTTGACACTACGTACACCACCTACTGCCCATTGTCAACCAC
TGAAGCTCCAGCTCCATCTACTGCTA
CTGATGTTTCTACCACCGTTGTCACCATCACCTCATGTGAAGAAGACAAATGTC
ATGAAACCGCTGTCACCACCGGTGTC

Fig 31

ACCACTGTCACTGAAGGTACTACCATCTACACTACCTACTGCCCATTGCCATCT
ACTGAAGCTCCAGGTCCAGCTCCATC
TACTGCTGAAGAATCTAAACCAGCTGAATCTTCCCCAGTTCCAACCACCGCTGC
TGAATCTTCCCCAGCTAAAACTACTG
CTGCTGAATCTTCCCCAGCTCAAGAAACCACTCCAAAGACCGTTGCTGCTGAAT
CTTCTTCAGCTGAAACTACTGCTCCA
GCTGTCTCTACCGCTGAAGCCGGTGCTGCTGCTAACGCTGTCCCAGTTGCTGCT
GGTTTGTTGGCTTTGGCTGCTTTGTT
TTAAGTTTATTAGAGCTTAAATCAAATATTTACAAACAAAATTTTCATTTTCCC
CCCTTTCCCTTTCTTCATTCTTCAAA
AAAGGGTTATTTACTATTAATTGATAAATTTATGGTTTCATGTTAATTTACCCTT
TTCTTTATAAACATTGGTATTATTA
TTATCATCATTAGNTTTATTTATATTTTCGTGAGTTTTTCGGNTTTAATTAATTTT
TTTGGATACATATTAAAAATTTAT
TTGGTACTAG

Fig 31 (cont)

>80g3 669bp in-house: 1-652 PathoSeq: 653-669

TTGCCAAAATTTTATAAAAAATTGTCAAATTGAAAAGAAGTATTTCCCAAAAT
AAATTGTTTTTTCATCACAACCGGTTC
ATATCGCCATAGNCCATTTTTAATCTTAAGGTTGATACCAGTTAATTGTTGATTT
CTCTGTTATAGCCCCTGTCTAAATC
TGTCTATTTCTGGTATCGAATCAAAATGTCGCTCATAATGTGCATGTCGCAAAG
ATGTCGTAAAGTTTTGATTTCATACT
CATCTTAAATTTTTTTTAGTGATTGGCATTTTGTTCTTTCACATAGTTTTTATTTC
TAGTTATCAACCTATCAAATACAC
CTCCACAACAATGCATCCAAATAATAAAAATTCATTTAAATCAAAAAAGAAAT
TTATAGATCGTCGAGAAGCCAAGTCTC
AAGATATAAAACGTGCATTAACCCATAGGGCTAGATTAAGAAAGAACTATTTC
AAACTATTAGAAAAAGAAGGGTTACAA
GAGGAGAGGAAGCCTGAAGATGAGAACGATATAAGACCAACCAAGAAGAAG
GGAATAAATTTTGAAGAACGTGCAGCCAT
TGTGAAACAACGTAAAGAGGAAAAACGTAAATTCAAACTAGCAAGTGTACAA
GCAAAATTGGAAAAGATTGAATCTAATT
CGAAAGAAAGAGCTTTAAACCGTGACCAC

Fig 32

>85g3 481bp in-house:1-431
CTAATATACGTCGAGTTCTGGGGGGTTGAAAAAACGGGTATTTTTTGGACCAGCAGAAAAAAAGTGGATTTGCGCGTGCA
CGACCCGAAAAAGGGAGAATTTTTGAAAAATGGCGAAATTTGGGGTAAGTTTGAGAGAGTGTGGAGCAACAACTATAAGA
GAGGGTGAGCGCAAATTGTAATGGCAGGTCGGCAGGCCAATGAAGATGTGTTGTGCAAAAGATGGAGTTTGTAGCGGTTG
CTGTGGCTGAGATATTGGCACTTTTTTAAGACCGCATGTTTTTGGGTAGCGCTGGGTTAAGACCACATTTTTTTTTGTTA
GAAGACCGCAGAAAAGAGAGCACACATACAAAATCAAGACCGCAGAAAAGAGAGCACACATTTAAGAGCACATTTTTGGT
AGCACACACTTTTAAGAGCACAGAAAAAAGAGCACTTATTTCTAAGACCGCATGTTTGGTAGCACACACTTTTAAGAGCA
C

Fig 33

>21g2 667bp in-house: 1-667

TTTAGTTTTATATTGATGATGTTTTTAAGTGCTTGTTTATCATGGTGGATGGAAA
TTAGAATGAGTAAATTGAATGGAAA
ATCACTGCAACACCAACAACAACCACTGGTGGATACGAAAATTTAGTGTACAA
ATTTCTGCCAAAAAAATACAATAAAAA
CCGCTTATAGTCTTCTACTGACATAACAACACAAGTCAATAAATCAACAACTC
ATAAACAATGTAGACTTAATACTATCG
CTTAATTATTTAAACTATAATAAATACCCTATAGTATTATGCCTTTGTCAATGTG
TGTAGAATTTGGTTATTACATATCC
ATGTGTNATATATATGTTGATCAAAAAAACGCGATCTTCTCTTTGGTGTAGTGT
GTTACNCAAAAAATTCACTANTCTAG
GTCNCTGANAATCACTTGAAAATCAAAAATTTGTTGAAATTGAATTTCCTCMA
YTTTGAAATTTTGTTTGAAATTTTTTT
TTTGCTTTACAAAAAGACTCCATTTTGTTTTCCATTTCACAACCAATTACTTAAT
TCCTCTTTTTCATAATTAATAACTA
TCATTACTTACAACTACAAACAACTACGATCATTTCCTAAGAAAAAGCAACGA
GGGCGAATTGAGACATTAATCCCCTTT
ATTTTATCATCATGCCTTATACAGAAC

Fig 34

>66g4 579bp in-house: 1-579

CCCCGTTAACCACTTCTAGGGTATACCATTTCATCTGACTGAATAACTGGTTAG
TCGATTTGTTGTTGAAGAAAAGTGAC
CACCTAGTTTTTTCTGCCAACATTTTTTGCGATGAGCCGTCGACGCGTTGTCTTT
TTCTACCCCACGTTTAACAATCTTG
CCAGTCAATTCCCTAGCCAAATAAACTTTAGACTCACAACTCTAACACTGACTC
GTGCCCCCCTGTTTAAACTCTAAATT
ACTTCACAGAGCCTTTACTACCTTAAATTTARGRTTWTSKAKKGTTTCTGTTTTT
TTGCAAATCACCCTGACTYGTTTTT
TTTTCAGCCAGGTTTTTCGTTAAAATCTGACCAAAAAATTTACRACTCCTATWT
TTAAAACTCYAAAWWACAATTAAAAC
TCAATTCAGACAAGTCCTTCTGCTCATTCTGAGTCTTCTCTATTGTCTTTTGACT
TTTTGTGTGTGACTATTTTCATGAT
CACCCCGTTTCTTGCATTTTTTTCAGTCAACTTTTTCTCAAAATCAAGCCAAAAA
AACACACCTTTAACTACCTATACAA
CGCAAACCTATTCAAAACA

Fig 35

Sequences with known function, C. albicans sequence NOT present in the public domain (ALCES/EMBL)

>CFL (223c_cp) 165bp in-house: 1-165

AACTATTGCCAATGGTAAATATGCCAGTGAAATCGAGAATTTTAATAAGTCGG
TCCCTCTTAAGGTCCCATTCAAATTCA
CTAATGCACAATTGGATCTTTATGCTGCTAGCACACATAACCAAGAGCCAATA
TCCTAGTAACGACGCACCATAGTAGAC
CGAAT

Fig 36

>EF4 (29g3) part1 479bp in-house: 130-479 PathoSeq: 1-129

CGCGAAGNNTCAATCATNTCAGAAGAAATGAAAGAAGGTACTCCGTTCTTTAC
TATTGTGGCAAGAATCCCTGTGATTGA
GGCATTTGGGTTTTCCGAGGATATTAGAAAGAAGACATCCGGGGCAGCTAGTC
CTCAATTAGTTTTTGATGGGTATGATA
TGTTAGATATCGATCCATTTTGGGTTCCACATACTGAAGAAGAATTAGAAGAAT
TGGGTGAATTTGCAGAAAGAGAAAAT
GTTGCTAGAAGATATATGAATAATATCAGAAGAAGAAAAGGGTTATTTGTTGA
TGAGAAAGTCGTCNAAAATGCTGAAAA
GCNAAGAACTTTGAAAAGAGATTAGATTATCCNGTTNAACAGGCCATATGTGT
GAAATTGTTTCCNAAAAGACAGATACN
ANGTGGNCCNTATTTGTTTAATATTCCACNACCAGTTAATGTTTTGATATNGAT
GTTTTATATAGTCCAATGTTGAGAC

Fig 37(a)

>EF4 (29g3) part2 1706bp

AAGTCATGCGATTGCAACAAGGATCACAAGAACCAGAAGTTCACGAACATTTG
ATTAATTTGATTGATTCACCTGGGCAT
ATTGACTTTTCGTCTGAAGTGAGTACTTCTTCGAGATTATGTGATGGTGCAGTT
GTTTTGGTCGATGTCGTCGAAGGTGT
CTGCTCACAAACAGTCAACGTTCTACGCCAATGTTGGATTGATAAGTTGAAGC
CATTACTAGTTATTAACAAAATTGATA
GGTTAATCACAGAATGGAAATTGTCTCCCTTGGAGGCATACCAACACATTTCC
AGAATTATAGAACAAGTAAACTCTGTG
ATTGGGTCATTTTTTGCTGGTGATAGACTAGAAGATGACTTGAATTGGCGTGAG
GCTGGTTCTGTCGGGGAGTTTATCGA
GAAGAGTGATGAAGACTTGTATTTCACACCTGAAAAGAATAATGTAATATTTG
CCTCGGCAATAGATGGATGGGCATTTT
CAGTCAATACATTTGCCAAAATATACCTGAAAAAATTAGGGTTCTCTCAACAA
GCATTGTCAAAAACTCTCTGGGGAGAC
TTTTACTTGGATATGAAAAATAAAAAAATCATCCCTGGTAAAAAATTGAAAAA
TAATAGTAACAGTTTGAAGCCATTATT
TGTTTCGTTGATTTTGGACCAGGTTTGGGCTGTTTATGAAAACTGTGTTATTGA
AAGAAATCAAGACAAGTTGGAAAAAA
TCATTGAGAAATTAGGGGCCAAAATCACCCCTCGTGATTTGCGATCCAAAGAT
TACAAGAACTTGCTAAACTTGATTATG

Fig 37(b)

TCTCAGTGGATTCCTTTGAGTCATGCCATATTGGGGTCAGTGATTGAATACTTG
CCAAGCCCCATTGTTGCTCAGCGTGA
AAGAATAGACAARAWWTWRRRKGRRMCSMYYTATARWRYWKTGKWTTCAR
AAMTGSATAWWTCCAAACTAGTCGAMMCTT
CATTTGKMAARRMKWTGCASRMAYSYSMKRGTWMACMCCCRKAAMCCMAT
WGGCMATWKYAKNTGTMYCAAAATTGTTTG
TCAATCCCCCAATGAAGACTTACCCAAAGCTAGTAATGCCCGCTACTGGAGGA
TTGACGGCCGATGAAATCCAAGAACGA
GGAAGAATTGCTCGAGAATTAGCCAAAAAGGCATCTGAAGCAGCTGCTTTGGC
ACAAAGAAGGTTCCCAAAAATGAAGAT
GAGTTTGCCATTAAACCCARGAAAGATCCATTTGAATGGGAATTTGAGGAGGA
CGATTTTGAGAATGAGGAAGATGAGAG
CGATGCAAACGCAGTTGAAGAATCAACTGAAACCATAGTGGGTTTCACTCGTA
TTTATTCTGGATCGTTATCTAGAGGCC
AAAAGCTCACGGTAATTGGACCCAAATACGACCCTTCATTACNTAGAGACCAT
CAAACCAACTTTGAACAAATAACCAGT
GAAGTGGAAATTAAAGATTTGTTTTATATCATGGGAGGAGAATTAGTGAGAAT
GGAAAATTTCCGTGCGGGTAATATTGT
TGGGGTTGTTGGATTGGATAACGCCGTGTTTAAGAATGCCACAATTTGCTCACC
GTAACGTGAAGATAAACCATACATTA
ATTTAGTTTCAACATCAACCTTGATCCACAATAAACCAATTATGAAAATAGCA
GTTGAACCAACAAACCCAATAAAACTA
GCAAAATTGGAACGAGGATTAGATTT

Fig 37(5)(cont)

>NDI (17c_cp) 807bp in-house: 1-614 PathoSeq: 615-807

AACCTATTCCATAATGTTTACTAGATCATTGATTAAAGGTGGTGGCAGACTTGC
TACTACCAGATCATTGGTCAACAACT
CTACTAGTTTGGTTTTAAAAAATCAATTTAAGAAATATTCAACATCAACTCCTC
CTAAGGTTGCCAAATCAAAATCTTCG
ACAATTGGTAAAATATTCAGATACACTTTTTACACTGCTGTGATATCGGTTATT
GGTTCTGCCGGTTTGATCGGTTACAA
AATTTACGAAGAGTCTCAACCTGTTGATCAAGTGAAACAAACACCATTGTTTCC
TAATGGTGAAAAAAAGAAAACTTTAG
TTATTTTGGGTTCTGGTTGGGGTGCTATTTCATTATTGAAAAACTTGGATACCA
CCTTGTATAATGTTGNTATTGTCTCC
CCAAGAAACTATTTCCTTTTCACCCCATTGTTACCATCTGTTCCTACCGGTACTG
TTGAATTGAGATCTATTATTGAACC
TGTCAGATCAGTCACCAGAAGATGCCCTGGCAAGTTATTTACCTTGAAGCAGA
AGCTACAAATATNAACCCCTAAAACTA
ATGAGTTGACACTTAACAAAGTACTACTGTCCGTTCTGGTCATTCTGGTAAAAA
TACTTCCTCTTCTAAATCAACTGTTG
CCGAATACACTGGGGTTGAAGAAATCACTACCACCTTGAATTATGACTATTTA
GTTGTTGGTGTTGGTGCTCAAACAATN
CTANTTTTCGGNAATCCTGGGAGNCGCNTGAGGAANTTCAACCCCTTTTTTGAA
AGAANGNCCAGTGGANGCCNTCTGCN
AATTAGA

Fig 38

>RPL27 (357cL) 560bp in-house: 1-560

AAAAATGGCTAAGTTCATCAAATCTGGTAAAGTTGCTATTGTTGTAAGAGGTC
GTTACGCTGGTAAAAAAGTAGTCATTG
TGAAACCACATGATGAAGGTACCAAATCTCACCCATTCCCACATGCCATTGTC
GCTGGTATTGAAAGAGCTCCATTGAAG
GTTACCAAGAAGATGGATGCTAAAAAAGTTACCAAAAGAACTAAAGTCAAGC
CATTTGTTAAATTAGTAAACTACAACCA
TTTAATGCCAACTAGATACTCATTGGATGTTGAATCATTCAAATCTGCTGTCAC
TTCTGAAGCTTTAGAAGAACCATCTC
AAAGAGAAGAAGCTAAAAAAAGTTGTCAAGAAGGCTTTTGAAGAAAAACATC
AAGCTGGTAAGAACAAATGGTTCTTCCA
AAAATTACACTTTTAAGAAAGGAACCACCTTTATTTGAATGTTTGTAATATAGG
TTGAATCAGAGAGACAAAGTAGAAGA
AAATACAAAAAAGAGAGTATATCTGTATAGTATAATTTAATGGGGGTCTAATT
TACTTACCACTTTATTCGTGCATTATT

Fig 39

>SADH (110c_af) 650bp in-house: 1-650

AACCTTTTGAAACGATTAANTNCAATCAAACAATCTTATTCAAAAGTACTCGC
AATACGTACAATGTCAATTCCATCTAC
TCAGTACGGATTTTTTTATAATAAAGCTAGTGGTCTTAATTTGAAAAAAGACTT
GCCGGTTAACAAGCCAGGTGCTGGTC
AATTGCTTTTAAAGGTTGATGCAGTTGGCCTTTGTCATTCAGATTTACATGTTCT
CTATGAAGGTTTGGATTGTGGTGAT
AATTATGTGATGGGCCACGAAATTGCTGGGACTGANGCTGAACTANGGTGAAG
AGGTGAGTGAGTTTGCAGATGGAGATC
GTGTCGCTTGNGTCNGNCCCCANTGGATGTGGNCTTTGCAAACACTGTCTTACT
GGTAACGATAAATGTNTGNACCAANT
CGTTTATTGGATTTGNTTTCGGATTGGNTTACAATGGNANGNTNCGANCCATTT
TTGGTAGNNANGAGANCNANAACTTG
GTAAAGATCCNTNTAATGTACNTNCCGAGNAAGCTGCNCTTTTNNNGNNTCCN
TATTGANTCNTACCAANGNTTTTAAGG
NATGNAGNAGTGTGNCANCNTGGGAATAATAATACACNTTCTTNCCTGATGGN
NTTNTGCNGTTACNCNNTTCAAGNNNN
CNNNAAGCNC

Fig 40

Sequences with known function, partial C. albicans sequence present in the public domain (ALCES)

>ABP1 (409c10) 1435bp in-house: 842-1435 public: 1-382/779-841 PathoSeq: 383-778

ATGGAAAAAATTGACATGAATACGTATTCAAACAATATCCAACAAGCATACGA
TAAAGTTGTTAGAGGAGAACCAAATGC
AACATTCGTCGTTTATTCTGTTGACAAAGACGCCACTATGGACGTCACTGAAAC
AGGGGACGGATCATTATACGATTTTG
TTGAACATTTTACTGATGGACAAGTTCAATTTGGTTTACCCAGGGTTACTGTTC
CAGGATCTGACGTCTCCAAGAACATC

Fig 41

TTGTTAGGATGGTGTCCTGACAGTGCTCCACCAAAATTGAGATTGTCATATGAC
AATAATTCTGCTGATGTGTCAAGAAT
ACTGAGCGGATACCATGTGCAAATTACTGCAATGGATCAAAATGATTTATACG
TGAATCACTTCTTGAATAGAGNTGGTG
CTGCTGCTGGTGCAAGATATTCCACTCAAACTTCCGGACTCAAAAAACCATCCC
CTGCTGCACCTAAACCTACTTCAAAA
CCTGTTGTTGCTAAATCTAGTTCTGCTTCAAAACCTTCATTTGTACCCAAATCTA
CTGGGAAGCCTGTTGCTCCAGCTAA
GCCAAAACCAAAGAACATCACCAAGGATGCTGGTTGGGGTGATGCTGAAGAC
GTTGAGGAAAGAGACTTTGACAAGAAAC
CTTTGGATAACGTTCCATCGGCATATAAACCAACAAAGGTTAACATTGACGAA
TTGAGAAAACAAAAATCAGATACAACT
AGCTCAACTCCTAAAACATTCAAATCTGAACCACAAGAAGAAAAGAATGACG
ATGATGGGCCATCCAAACCTTTATCGGA
AAGGATGAAAGCCTATGATCACGACTCAAGTCGTGATGGAAGATTGACTTCTT
TACCAAAACCAACGATTGGACATTCTG
TNGCCGATAAATATAAAGCTAGTGCATCTGGGAATGGTGCTGCTCCTGCGTTTG
GTGCTAAACCAGCATTTGGTCACAAT
CAGTTGATTCAAGAAAGGATAAATTGGTAGGTGGTTTGTCGAGAGATTTTGGT
GCTGAAAATGGAAAAACTCCGGCACAA
ATTTGGGCTGAAAAAAGGGGAAAATACAAAACAGTGGCCTCCGATGAGAAAG
AAACTAACTCAAGTGAAAAAGTTGATGA
GCCAGAGGAACATCATGCTGCCGACTTGGCCAAAAAATTTGAAGAAAAGGCA
AATATTGCTGGCGATACTCCTTCCTTGC
CAACTAGAAACTTACCACCAGCACCACCAGCACGAGAAACCGCAATTCCATCT
AACGAAAAAGACAAAAAAGAAAAGGAA
GAGGAAGAACAAGCTTCAGCACCATCTTTGCTACTAGAAACTTACCACCACCG
TNACAAAGACAACCTGAGCCCGACCAG
AACCAGANGAAGAGGAGGAAGAAGAAGAAGAGGAGGCTCCTGNTTCAAGCTT
ACCAGCAAGAAATCTCCCCCCAC

Fig 41 (cont)

>ADE12 (226c_af2) 993bp in-house: 1-646 public: 647-677 PathoSeq: 678-993

NATAAGGAATGANCCAAAGNAGTGTANNAGNTAATAAGNTNAGANANTTCCA
AAAGAANNAAAANAACCTAACNACANAN
NNNTATTANATCCAGTAGAAANACAANATTAAGNNACCANTATCTTNNAAAG
NTTNACGANACNTNGTTCNAATTGTTCA
TNTTGGGANGNACCCTCAACAACTNTANTNGGGTAAACTTTAACNACACCATA
GACATTTNTNATGGTTNTTGAGGAATA
CCCAACCCAGTCANAAANACCAACAATACCAGTTGATAATAAGTGANGTATGA
TAAGTACAGANATCAATATCCAACATT
AAACGCATTAGCACCTTCANCCAAGATTTTTTTATTGGCAGCNATAGCTTCGTG
CATNAAGTTGACGGAGTCTGACGACG
AATGGTCTCAAGGTTTCACGGNATTTTTCAAATCTTGCCAATTCTTCCTTAGGA
TCATATTCAAATTCACCGTATCTTTT
TTGTCTACTCTCGACTAATCTCAAATATCTAGTTTTGAATTCTTCCCAAGCTTCT
GGATCAGGGTTGACTAAATGGTGGA
CTCTGATACCTGATCTACTTGCCTTGGTTGAGTAAGTTGGACCAATACCTTTAC
CGGTAGTACCTATTGATTTCTTATTG

Fig 42

GNTGTTAATTCAGCATCTTTCAATTTATCAGCACGTTGATGGAAGTCAAAGACC
AAATGAGCTCTAGATGAAACAAACAA
TCTATCACGACAATCTAACCCTTTTGCTTCCAAGTTTTCCAATTCAGCAAAGAA
GGAAGGAACGTGGATAACAACACCAG
ATCCAACTAAGTTTTGACATTTAGGATTGACCAAACCAGAAGGTAACATGTGG
AAGTCATACTTGACTTTACCAACAACA
ATCGTGTGGCCAGCATTGTTACCACCTTGACATCTGGCACAAACATCGATATCA
TCACATAATAAATCGACTAATTTACC
TTTACCTTCATCCCCCCATTGAGATCCTAATAC

Fig 42 (cont)

>CDC48 2448bp in-house: 95-220/285-1340 public: 1-94/221-284/1341-1373/1783-2273

PathoSeq: 1374-1782/2274-2448
CGTTTCCGAWAGCTGCACCTGAGTTGGCACCTGCTGCTGAACCATTATCAGTG
GCACCAGCATTTTCATTGAATCTAAAG
CTAGAAAATTGACCTCTTGAGGCTTGCAATTGTTGAGCGTAAGACTCATAACG
ACGTAATTCAGCGTCTGAAACAGATCT
TTTTGCGGTCTTCATAGCCTCTTCAAAGTGAGCTCTGGTAATGTAAGGCACAGG
GTCTTCTTCTTCAACTTCATCTACCT
TCATATCAACATCTTCAGTTATCACCTGTTCCTTTTGCTTCTTTAATCTTGTTAAT
CTTTACTTGGGCTTCAATAGAGTC
TTTAATAGCAAATTTAGCAGATCTTTGAACAATATAAGACAAATCTGCACCCG
AGAAACCGTGAGTGATCTTGGCAATTT
CGTTCAAGTCCAAACCAGGTTCTAATGGAGTGTTTCTCAATTGAGCTTGTAAAA
TAGACAATCTAGCTGGCTCATCTGGC
AATGGGACATAAATTAATTGATCCAATCTACCTGGTCTCAATAATGCAGGATC
AATTTGATCTGGTCTGTTAGTGGCACC
AATGACAAACACATTCTTCTTAGCATTCATACCGTCCATTTCAGTCAACAATTG
ATTGACCACTCTGTCGGAGGCACCAC
CAGCATCACCGTGAGAACCACCTCTAGCTTTGGCAATGGAGTCCAATTCATCC
AAAAACACCACAGTAGGAGCAGCAGCT
CTGGCCTTGTCAAATATATCACGGATATTAGACTCAGATTCACCATACCACATA
CTCAACAATTCTGGACCTTTGACAGA
AATGAAATTAGCAGAAACTTCAGTAGCAACAGCCTTGGCCAAAAGTGTCTTAC
CAGTACCTGGTGGACCAAAGAACAAAA
CACCTTTTGTTGGTGCCAATCCGAATTTTTGGTATTGATCTGGATGTAAAACAG
GATACTCCACGGTTTCTTTTAATTCA
TTCTTAATGTTGTCCAAACCACCAATATCATCCCAAGTGACATTAACATTTTCA
ACAACAGTTTCACGCAAGGCAGATGG
GTTGGAGTTTCCGAGAGCAAATCTGAAGTTGTCTTGAGTGACACCCAAAGAGT
TCAACACTTCAGTATCAATGGTTTCTT
CTTCCAAGTCGATAAGATCCATCTTTTCACGGATTTGTTGCATAGCAGCTTCTG
AACATAATGAAGCAATATCAGCACCA
ACGAAACCATGTGTTTCAGAAGCGATGGCTTCCAAGTCAACATCATCAGCCAA
TTTCATATTCTTTGTGTGGATTCTCAA
AATCTCTAAACGTCCTTCAGCATCCGGAACACCAATGTCAACTTCTCTGTCGAA
TCTTCCATCAATAGAATTTTGAAATC
TTCTCAAAGCAGGGTCCTGTTAGTAGCAGCCAATTAACAACCTACATTAGATCT
GGCCCTTCATACCATCCATAAGGGGT

Fig 43

TAACAATGGAGAAACAACTCTTCTTTCTACTTCACCCATTAGTTTTGTCTCTCTT
TGGGGCAATAGAGTCAATCTCATCA
ATGAAAATAATGGAAGGAGAATTCTTTTCAGCCTCTTCAAAAGCTTTTCTTAAA
TTGGATTCAGACTCACCAGCCATTTT
AGACATAATTTCTGGACCATTTATTAAGAAAAAGAAGGCACCTGTTTCATTGG
CCACTGCTCTTGCCATAATGGTTTTAC
CGGTACCAGGTGGACCATACATCAAAATACCCTTTGGTGGCTTAATACCAATC
GATTTGAATAATTGTGGATGTCTTAAA
GGCAATTCAACCAATTCTCTAATTTGGGCCATTGTTTCTTACACCTCCCAATAT
CGTCGTAACCCACTTCATTCAAGCTA
TTTTCTTCATCTTCACGATTAATAGGTTCTCCTTCACAATGAATAATGGTATCTT
GAGCAACAATTGCAATTTCTTCAGG
GTCAACTTCAACAACTTTGAATTCTACTTGTCTCATACCACCCCTCACAGTGAA
TAAATCACCTTTTCTCACTGGTCTAT
AGGCTTCAACAAAATATGGCTTCAAGTAAAGGTCGAATAAGGAACCATTAATA
CCTTCAACAGTATCAGCAATTGGCAAT
ACTGAGATTCTGTTGGCATATTTAATATCAGGACATGGATGGACAGTAACGAT
ATCTCCCAATCTGACACGCAAATTGTT
ACGAACACATCTGTTAACTCTAGCAACGCCATCAGGCATATCATCATCAGCTA
AAACGATCAACACTGTGTCCTTTCTCT
TCTTACCTTTCACCAAGACTGTATCACCACGGAATAATTGTAACAATTCCATTG
TGTTTGACGACATGGTTATGACAGAA
TTGTCATCGTTGGTGGCGTCATCAACGACCAAGGCATTATCTTTCTTTGGTCTTC
TTAAAATTGCAGTTGCTGTTTTATC
ATCTACAGCAGAAGCACCAGAAGCATCAAAATGTTGTTTTTTATCTTC

Fig 43 (cont)

>CIT (99g3) part1 1435bp in-house: 803-1435 public: 1-333 PathoSeq: 334-802

TCCAAAACTTATTGCTTAGCTATACGGTGTAATGGACCAGCTTAACCATTCAAA
CCAGCAGCTAATGACAAGAATGGAGA
AGATAAAGCGGTACCAACCAAGTGGGTGGTGTGTGCAGAGACGTTACCACCTT
CGTGGTCAGAGTGGATGGTAAGGTACA
ATCTCATTAATTCAACANATTCCTTGGTGTCACCAAAACCTAACANACTGGCCA
AGTTAGCACCGTAATCCAATTTGGAG
TCAATGGCAGCTGGCAATTTACCATCGTGGAAAACGTTTCTGTAAATCTTAGCA
GCAATGGTTGGCAATTTAGCTAACAA
ATCGATGGAATCTTCGTAAGTGTATTTCCAGTATTCGGATTTGTTGGCACCTTT
AGCATAAGCTTGGGCAAATTGGGATT
CAGATTCCAAAGCAGTAACGGCAATGGAGAATTGAGCCATTGGGTGCAAGTGA
GATGGAGATCTGTCGATCAATTCTTCA
ACGTGCTTTGGTAATGCTGATCTAGCAGCAAATTCTTCGGATAAAGCCTTAGTT
TGGGCGTCAGTTGGAACTTCACCAGT
CAACAACAACCAGAAAAGAGCTTCTGGTAATGGTTCTTCACCACCTGGTGCTTT
TGGCAATTCTTTTTGAATGTCTGGGA
TGGTTCTTCCTCTGAAACGGATACCTTCAATTGGGTCCAAAACAGAACCTTCCC
AAACTAAACCTTTGATACCTCTCATA
CCACCGTAAGCTTGTTCTAATAAAACTTCACCCAATGACAGTTTTACCGTGTTC
TTTTTTGAATTGTTTAACTTCTTCAG

Fig 44(a)

CTTTGGCTGGCAAGATTTCTTCCAATCTTTGTTTTAAGGTCTAAGAAAGCAAGT
TAGTAAATTATCTTTGTATATACGTT
AAAAGTAAAATCCCAAGCACATTGCCCGGATCCTCAAAAAAGTGAATACATAC
TGGTTCAGCAGAAGCATATGTTCTGAT
GCTGTTTTTGAAAGTGCTCTTGGCTACGTTGGTTGAACGTTGAATTGATCTGAA
TGCAGACATTGTTTGATAAATATACT
ATATTCTAAAAGAAACTTAAAAGATGGAAAAAGGAAAAGAAGAAATGGAAAA
AAAAAAGTAAAAGGAAATCAATTGCAAA
TATATACAAAAAAATCCAGCANGAAATCAGTTGAAATTTATATTCCAAATTTTT
GGTTTTAATGGCTCTTGAAGTTGTGG
TGAACAATTTTTTTTTTTTTACTTTTTTCCTCATGGATTTACTTTAGTTTTGGGTC
TGTCGNGCTGCCGTACAACTTTCC
TGNGAAAATTGATTTTTTTTTCTTCTGGNGAGGATTTTTTGGCGTTCTTTGTTAA
ACTTCTTATTTGTCCCACTANATGG
AGAGCAAAAAAAAAAAGTTTGACTTTTACTTTTAACCAATCAATTTTCCAGAATN
TGAACGAGAAAAAGGGAAAAN

Fig 44a(cont)

>CIT (99g3) part2 327bp public: 1-327

CAAAGAGAATTTGCTCTTAAACATATGCCAGACTACGAATTGTTCAAATTGGTT
TCAAACATTTACGAAGTCGCTCCAGG
TGTTTTGACCAAACACGGTAAGACCAAGAACCCATGGCCAAATGTGGACTCCC
ACTCTGGTGTCTTGTTACAATACTACG
GTTTGACTGAACAATCTTTCTACACTGTCTTGTTCGGTGTTTCCAGAGCCTTTGG
TGTCTTGCCACAATTGATCTTGGAC
CGTGGTATCGGTATGCCAATTGAAAGACCAAAATCTTTCTCCACTGAAAAATA
CATTGAATTGGTCAAAAACATCAACAA
AGCTTAA

Fig 44 s

>HOL1 (409c5) part1 695bp in-house: 98-695 public: 1-97

TTTCTGGATTATCATGTTATTTGGTTAGCTANACGGAATAATGGGATAATGGAA
GCTGAATATCGATTATATNTATTAGT
TATCACTTTAATCATTTAACCCGTAGGGTTAATTATGTTTGGTGTTGGTGCCGCT
AGAGAATGGCCATGGCAAGTGATTT
ATGTTGGATTAGGTTTCATTGGGTTTGGTTGGGGATCAATTGGTGATACTTCAA
TGTCTTATTTAATGGATGCTTATCCT
GATATTGTCATTCAAGGAATGGTGGGAGTAAGTATTATTAATAATACTTTGGCT
TGTATTTTCACTTTTGCTTGTTCTTA
TTGGTTAGATGGATCAGGAACACAAAACACATATATTGCCTTGTCAATTATTGA
TTTTGCTACCATAGCATTGGTTTTCC
CCTTTTTATATTATGGTAAAACATTTAGAAGGAAAACTAAAAGACTTTATGTTT
CAATGGTTGAATTGACTCAAGGGATG
GGATAAGAGAGTGAGTGGTAAAAGAATTTTATTAATGATACATTTATTATTAG
AATTACTACTATGGAAATCCGAGTCTG

Fig 45(a)

TGTTTTTTTTAGAAGTATATTTTAGACGTATTTAGAGTTGGTTTTCTCCTTTGTA
CTNTATTTAGCATTTTATAATATAT
TAAATTCAAGTTGCATTAATATATATAAATAAAAAAACCTACNAANAAAAAAN
GN

Fig 45a (Cont)

>HOL1 (409c5) part2 762bp PathoSeq: 1-762

GATCAGAATAATGAGGACTTTATACCTGGAACACTCAATATCTATTCCTTGGAA
GTTGACTCTGAAGATGAAAACGTGAG
TCATTACGATGCTTCCAGTCGACCAAAAGTGAAAACAAAAGGCAATATAATCC
TCTTCCCACAACCATCGAATTCATGCA
ATGATCCATTAAATTGGAGTAAATGGAGAAAGCTAAGTAACTTTTTTATTGTCA
TTTTTATTACTGCTTTTACAGCAGCT
ACTTCAAATGACGCTGGATCAATTCAAGATTCACTTAATGAAAAATATGGAAT
TAGTTACGACGCAATGAATACAGGGGC
AGGCGTTTTATTTTTGGGTATTGGATGGGGTACTTTCTTTTTAACACCTGCTTCG
TCGTTATATGGTCGAAAAATAACAT
ACTTTATATGTATCTTTCTTGGTTTATTAGGCGCTGTTTGGTTTGCCTTGGTTAA
AAGCACTTCCGACTCAATTTGGTCG
CAATTGTTTGTTGGTATTAGTGAGAGTTGTGCTGAAGCTCAAGTACAATTAAGT
TTATCAGAACTTTATTTTGCCCATAA
CCTTGGTTCTGTGCTTACGTCCTATATTGTTGCAACTTCCGTAGGTACTTACTTA
GGACCTTTAATTGCAGCCTTTATTG
TTCAAAACATTGGTTTTAGATGGGTTGGTTGGATTGCAGCAATTATTAGTGGTG
CATTATTGTTCGTAATTGTTTTTTGT
TTAGATGAAACCTATTTTGATCGAGCAAAGTTTACCAAGCCA

Fig 45b

>ESP1 1458bp in-house: 889-1458 PathoSeq: 1-888

CGATTTTCAATTACAAGATATTTTGCATCATGTTGAAAGCAAATGGTTTGGTGG
GTTTATTTCAGGTATTTTCACTAATG
ACAATGACGTTGAAAATGAATCCAAGAACGTGTTTCATAAATTCAAACAAGAT
TTAATGAAAATTTTGAAAGATTGTTTA
ACCGTAAGTGACGATAAATCGAATATAGAGAGGTTTCTTCAGTTTAATGAATTT
ATTTATTACTGCTTTTACTCAATGGA
GGAATATAATTATGAATTGGTTGATGATTTGATAAAATTTATAACTATAAATAT
GAATTCTCATGGCAGAATAGTTAATT
TTGGCACTAATGTTAAAATTAATANATTACACGAATTAATTAAGAATTTGATTG
ATAAAGTTAATAAAAACAAACAAAGA
TGTGACTAGCAACAACAAAAACAACAGCAACAACAACAGCAACAACAACAGC
AACAGCAACAATTCCCAACATATTGGTT
TTGATACCTAATGCCAACTGNTCCAATTTCCCAATGGGAAATCGATGGAANTTC
NTTCGTAAGTAAAATCCAATTTCAAG
GAATNGCAATCAANTCCTTANGTTACTTGATCTAGTCAAATCAAACACCAATA
ACAAGAACAAGTTAATGTTTGTTGATA
AATCTAATTTGTATTATTTGATTAATCCCAGTGGTGATTTAATTCGATCAGAAA
ATCGATTCAAAAAACTATTTGAATCA

Fig 46

AATCATTTATGGAGAGGGGAAATTGGAAAATTATCAAGTAATGAACATGAAGA
TTATCAAGATTCAATATTATGTGAAAT
CTTGAAAAGTCATTTATTTGTTTATATTGGTCATGGTGGTTGTGATCAATATATT
ANAGTATCAAAATTATTTAAAAAAT
TACCTCCTAGTTTATTGTTAGGTTGTTCATCAGTTAAATTAGATAATTGTAATTA
TAACTATAATTCCAGTATGTTACAA
CCACTGGGTAATATTTATAATTGGTTGAACTGTAAATCGTCAATGATACTCGGG
AATCTATGGGATGTTACTGATAAAGA
CATTGATATTTTTACACTTTCATTACTACAAAAATGGGGGTTAATAGATGATTA
TAATGGCAGTGGCCATGATTATGGTA
TGAAGAAATTGGATTTGACTAATTGTGTTGTTCAAAGTCGAAGTAAATGTACTT
TGAAATACTTGAATGGATCAGCACCT
GTGGTTTATGGTCTACCAATGTATTTAAAATAGACATTCTGTTTGCATATAAGT
TTATATATTTTAATAATAAGAAAAAG
AGCATAATTTGGATCTTGATTTTGTATTGTTTGGTTTGTTATGAACAAATTTTGC
ACCCAATCACTATCGAACTTTCTTT
TTTAAACAGAGAACATTTAATCAACATTTATGTTACATTTAAGCGTTTAAATAC
ATATTTGTGTTAGATAGTTATATAAT
GTTTGATGCAAACATACA

Fig 46 (cont)

>FAL1 (190g3) 1439bp in-house: 1-770 public: 861-1299 PathoSeq: 771-860/1300-1439

CTTCTTTTAGAGACAATGCAGTGGTTTTCTTACCAGATGCATGACCCCCACCCA
ATAAAACTATAATCGATCTATTCACA
GTATTTGATGCCATTTTGATGGTGATGAATGATGTGATGTGATGCTCATCTTAT
TGGGAGTTTCAAAAAAAAAAGTTACA
CTCGAAAAAAAAAAAAATAGCATTATAAATAGAAGCTTTACTATCTTATAGAAC
AAAACAAAAAACACTATCTTCTAATTA
ATAATGGATGATTTTGATAGAGATTTAGATAATGAGTTGGAATTTAGTCATAAA
TCAACGAAAGGAATAAAGGTTCATCG
CACTTTTGAAAGTATGAATTTGAAACCTGATCTTTTGAAAGGAATATATGCCTA
TGGATTTGAAGCACCATCTGCTATTC
AATCTAGGGCTATTATGCAGATCATCAGTGGTAGAGACACAATAGCACAGGCA
CAATCTGGAACTGGTAAAACTGCTACT
TTTTCTATTGGTATGCTTGAGGTTATAGATACTAAATCAAAAGAGTGTCAAGCA
CTTATCTTGTCTCCTACTAGAGAGTT
GGCAATTCAAATACAAAATGTGGTCATGCATTTAGGAGATTATATGAACATTC
ACACCCATGCCTGTATTGGTGGGAAAA
ATGTCGGTGAGGATGTTAAGAAATTGCAGCAAGGGCAACAAATAGTTAGTGGG
ACACCAGGTAGAGTGATTGATGTGATA
AAAAGAAGAAATCTACAAACTAGAAATATCAAGGTTCTTATTTTAGATGAAGC
TGATGAACTTTTTACAAAAGGGTTTAA
AGAACAGATCTACGAAATCTACAAACATTTACCACCTTCGGTTCAAGTAGTAG
TTGTTAGTGCCACATTGCCACGTGAAG
TATTGGAGATGACAAGTAGTTTACCACTGATCCAGTGAAAATCTTGGTGAAGA
GGGATGAGATTTCGCTTCTGGGAATCA
CACAATATTATGTTCAATGTGAACGTGAAGATTGGAAGTTTGATACACTATGTG
ATTTGTATGACAACCTTACAATAACT

Fig 47

CAAGCAGTGATATTTTGTAATACCAAATTGAAGGTGAATTGGCTTGCTGATCAA
ATGAAAAAGCAAAACTTTACTGTTGT
GGCAATGCATGGTGATATGAAACAAGATGAACGAGATTCAATTATGAACGATT
TTAGAAGGGGGAATTCAAGAGTATTAA
TATCTACAGATGTTTGGGCAAGAGGTATTGATGTCCAACAAGTCTCGTTGGTAA
TAAATTATGATTTGCCCACCGATAAG
GAAAACTATATTCATAGAATTGGACGATCAGGTAGATTTGGTAGAAAGGGAAC
AGCTATAAACTTGATAACTAAAGATGA
TGTGGTCACTTTAAAAGAATTGGAGAAATATTATTCAACGANAATTAAGGAAA
TGCCAATGAATATTAATGATATAATG

Fig 47 (cont)

>FBP1 (40c_af) 638bp

AACGTTGGCCTGGCCCAGTTAATTCCGTTTCCAAGCAAATGAATGTCGATACCG
ACATCATCACGTTGACCCGTTTTATT
TTACAAGAACAGCAAACTGTTGCTCCCACCGCCACCGGTGAGTTGTCGTTGTTG
TTGAATGCGCTTCAATTTGCATTCAA
GTTTATTGCCCACAATATCAGAAGAGCTGAGTTGGTCAACCTTATTGGTGTTTC
TGGCTCTGCCAACTCTACCGGTGATG
TTCAGAAGAAATTGGATGTGATTGGTGATGAGATCTTTATCAATGCCATGAGAT
CTTCCAACAACGTCAAGGTTTTGGTT
TCTGAAGAGCAAGAAGACCTTATTGTGTTCCCAGGTGGTGGCACATATGCTGTT
TGTACTGATCCAATTGATGGGTCGTC
CAATATCGATGCTGGTGTTTCTGTTGGTACGATTTTTGGTGTGTACAAGTTGCA
AGAGGGGTCTACTGGTGGCATCAGCG
ATGTCTTGCGTCCTGGTAAGGAGATGGTCGCTGCGGGGTACACCATGTACGGT
GCATCTGCCCATTTGGCATTGACTACA
GGTCACGGNGTCAATCTTTTTACTTTGGATCTCANATGGGTGAATTTATTTTGC
CNATCCAAACTTGGAAGTTCCAGA

Fig 48

>GAL2 (360c6) 1004bp in-house: 625-1004 PathoSeq: 1-624

TCCATTTTCCCTTTTCCTCTTTTTCTACATCATCCTCACANCAATTTCAAATATG
TCTCAAGACAACGTCTCATCAACAT
CTACAGCTGAGGCTGTAAATAATGAAATCAAAGTCAAAGATGAATTTCCACAA
GAAGAACAAGCTCATACTAGTTTAGAA
GATAAACCAGTGAGTGCATACATTGGTATCATCATTATGTGTTTCCTTATTGCC
TTTGGTGGTTTTGTTTTCGGTTTCGA
TACTGGTACCATTTCTGGTTTTATTAATATGTCTGACTTTTTAGAAAGATTCGGT
GGTACTAAAGCTGACGGTACTCTTT
ACTTTTCCAATGTCAGAACTGGTTTAATGATTGGTTTGTTCAACGCTGGTTGTG
CCATTGGTGMWTTATYCTTGTCYAAA
GTCGGTGATATGTATGGTAGAAGAGTTGGTATCATGACTGCTATGATTGYCTAT
ATTGTTGGTATTATTGTTCAAATTGC
TTCTCAACATGCTTGGTATCAAGTCATGATTGGTAGAATTATYACTGGTCTTGC
CGTYGGTATGTTATCAGTTTTATGTC

Fig 49

CTTTGTTCATTTCCGAGGTTTCTCCAAAACATTTGAGAGGTACTTTGGTGTGCTG
TTTCCAATTGATGATTACCTTGGGT
ATCTTCNTGGGNTATTGGCTACCTATGGTACTAAGAGTTACTCAGACTCTAGAC
AATGGAGAATTCCATTAGGTTTATGT
TTCGCCTGGGCTTTATGTTTGGTTGCTGGTATGGTTAGAATGCCAGAATCTCCA
CGTTACCTTGTCGGTAAAGACAGAAT
TGAAGATGCTAAAATGTCACTTGCCAAAACTAACAAGGTTTCTCCAGAGGACC
CAGCATTATACCGTGAACTTCAATTAA
TCCAAGCTGGTGTTGAAAGAGAAAGATTGGCCGGTAAAGCATCTTGGGGTACT
TTATTCAATGGTAAACCAAGAATCTTT
GAAAGAGTTATTGTTGGTGTCATGTTACAAGCCTTACAACAATT

Fig 49 (cont)

>KGD2 (98c_cp) 334bp in-house: 139-334 public: 1-138

TTCTAACAACAACATCTTTCTTGGATCTTCAATCAATTCCTTGATGGTTCTTAAG
AAAATAACAGCTTCACGACCGTCAA
CTACTCTGTGGTCGTAAGTCAATGCTAAGTACATCATTGGTCTAGAAACGATTT
GTCCGTTAACAGNAATTGGTCTTTNT
TTAAAANTGTGTAAACCAAATACGGNAGTTTAANGCATTTTTATAATTGGGGT
ACAGTATAATGATCCAATAACACNGNC
ATTANAAATAGTGAAAGAACCNCCGGTCATATCTTACAAAGTCAATTTACNAT
TTCTGGCTTTNTTACNCAAATTANANA
TTTCCTTTTNAATA

Fig 50

>MAA (249c_af) 619bp

AACCCCACCTTCAAAGACAAAGAAGATTTCGTCAAGCAAACGAATGTCAGAGC
AGAAAAGAACCAAGAACTAATCAAATT
TGCCCGTGACAACCTTAACCATTTACCATTCACCGAAAAAGACGGAGGTGCAT
GGGAAAACTATGAACGAATGATCAGTG
GTATGCTCTACAACTGTTTACAAAAAGAATTGGAAACAACACGTATGTCTTGC
AGAGACTACATGTTGGACTACGGCAGT
TTCAGAACTAGAGATTATAAAACAACCCAAGAATTTCTTGATGCAAAATACAA
ACATTTAGAAAGTTTCATTGGACATGT
TGGCAAAAATGCATTTATGGAATATCCAATCTATTTTGATTATGGGTTTAACAC
TTATTTGGGTGATAATTTCTATTCCA
ATTACAATTTGACAATTTTGGATGTTTCCATAGTCAGAATTGGTAATAATGTCA
AGTGTGGTCCCAATGTATCTATCCTT
ACCCCAACACACCCAGTGGATCCCACTTTGCGCTATGATCAATTGGAAAATGC
CTTGCCTGTGACGGTGGGTAACGGGGT
CTGGTTGTGTGGAAGCTGTACCATTCTTTGGTGGGGTGACAGTANGTGATGGCA
GCATT

Fig 51

>MEG1 (55g1) 1380bp in-house: 1-368 public: 497-1096 PathoSeq: 369-496/1097-1380

AATTACAATCTGGTTTGTTACTACCATATCCCATTAGTGTTATTGTCATTGTAGA
TATTGATAATGGTTAAAGGATTGGT
TTTCATTTTTTGTGTAATGAATGAGCCAAAATAAAAAATCAATTCGATGCGATG
CAATGAAGTTTAATAAAATTTTTTTT
TTTCTTTATTTCTTTTAATCAACCCNNCNATCNTTAAATTGAATCAATACCNACC
ATTAACATACTTCTATATNCNTATA
TTTNTTTTACAAAATATCNTGGGGNAGANAACAACTAGTGNTNCNAAAACAAA
ACNACNTCCTTATCCNTTATTAAAANA
NATTTCNTCCCCAGGNGGGNATTTAAGAACCGTTCCCAGANNATCATCATCAT
CATCATCACAAAAGAAGAAATCATCAA
AGAAACAAAGACATAACGATGAAGACGACGAAGAAAATGGTGGCGGTGAAGG
ATTTTTAGATGCTTCTAGTTCAAGAAAG
ATTTTACAATTGGCAAAAGAACAACAGGATGAATTGGAACAGGAAGATGAAA
CACAAAATCACCTTTCATTTGTTCAATC
ATTAAAAAAATCAACAAATAGATAGTGAAGAAGAAGAAGAGGAAGATGATTAT
TCAGATTTGGAAGAAGAAGAAGAAGTTG
AAGAGATATTATATGATGAAGAAGATGCAGAATTTATTCCCAAAGATGCAGAA
TTATTTAATAAATATTTCCAATCCAGC
GGTGAAGATAATAATAATGATGATGATAATTCATTTCAACCAACAATAAATTT
AGCTGATAAATTTTNAGCCAAAATTCG
AGAAANAGAATCCCCACAACAACNACAACAACAGAGTTTTCCAGATAATAGT
AATGAAGATGCCGTATTGTTACCACCAA
AAGTCATTTTAGCTTATGAAAAAATTGGTCAAATTTTATCAACTTATATTCATG
GGAAATTACCTAAATTATTTAAAATT
TTACCAAGTTTAAAAAATTGGCAAGATGTATTATACGTGACAAATCCAAATAG
TTGGACTCCTCATGCCACATATGAAGC
AACTAAATTATTTGTGTCGAATTTATCAAGTAATGAAGCTACAGTTTTCATTGA
AACTATCTTGTTGCCACGATTCCGTG
ATTCTATTGAAAATTCCGATGATCATTCATTAAATTATCATATTTATCGAGCATT
AAAAAAATCATTATATAAACCAGGA
GCTTTTTTCAAAGGGTTCTTGTTACCTTTAGTCGATGGTTATTGTTCTGTACGTG
AAGCCACTATTGCTGCTTCAGTGTT
AACTAAAGTTTCTGTCCCTGTTTTACATTCATGTCATTATTGTGGCGTACTGATG
AATAAAAAACGAGAATCACCTGTAT
TTGTCCTACGGCGAATATAA

Fig 52

>RNR1 (38) 2562bp in-house: 1-2562

ATGTATGTTTATAAGAGAGATGGCCGTAAAGAGCCAGTACGTTTCGACAAAAT
CACTGCCAGAGTTCAAAGATTATGTTA
CGGTTTGAATCCAAACCACGTTGAACCAGTTGCTATTACCCAAAAAGTTATATC
AGGTGTTTACCAGGGGGTTACTACTA
TTGAGTTGGACAACTTGGCTGCAGAAATTGCTGCTACAATGACAACAATTCAC
CCAGATTACGCTGTCTTAGCCGCTAGA
ATTGCCGTATCAAATTTACATAAGCAAACCACCAAACAGTATTCCAAAGTGTC
TAAGGATTTATATGAATACATTAATCC
TAAGACTGGGTTACACTCTCCTATGATTTCCAAGGAAACCTACGACATCATTAT
GGAACACGAAGATGAATTAAACTCAG

Fig 53

CCATTGTTTACGACAGAGATTTTAACTACAATTATTTTGGGTTCAAGACTTTGG
AAAGATCATATTTGTTACGTATCAAC
GGTAAGGTTGCTGAAAGACCACAACATTTGATCATGAGGGTTGCTGTCGGTAT
TCACGGTAATGATATACCAAGGGTCAT
TGAAACCTATAACTTGATGTCTCAAAGATTCTTCACCCATGGTTCTCCTTGTTTA
TTTAACGCTGGTACACCAAGACCAC
AAATGTCCTCATGTTTCTTGCTTGCTATGAAGGATGATTCTATTGAAGGTATTT
ACGACACTTTGAAATCGTGTGCTTTG
ATCTCAAAAAGTGCTGGAGGAATCGGTTTACACATCCACAACATTCGTTCTACC
GGTGCTTACATTGCTGGTACCAATGG
TACTTCTAATGGTATTATTCCAATGGTAAGAGTATTCAATAACACTGCACGTTA
TGTCGACCAAGGTGGTAACAAGAGAC
CTGGTGCCTTTGCCTTGTACTTAGAACCATGGCACAGTGACATTTTTGATTTCA
TTGATATTAGAAAGAATCACGGTAAA
GAAGAAATCAGAGCCAGAGATTTGTTCCCAGCTTTGTGGATTCCAGATTTGTTC
ATGAAAAGAGTTGAACAAAATGGTGA
CTGGACTTTATTCTCACCAAATGAGGCCCCAGGCTTGGCTGATGTTTATGGTGA
CGAATTCGAAGAATTATACACCAAAT
ACGAAAAAGAAAACCGTGGTAGACAGACCATCAAAGCTCAAAAATTGTGGTA
TGCTATTTTGGGAGCCCAAACTGAAACA
GGTACCCCATTTATGTTATATAAAGATTCATGTAACAACAAATCCAACCAAAA
GAACTTGGGTATTATCAAATCTTCCAA
CTTGTGTTGTGAAATTGTTGAATATTCTGCTCCAGATGAAGTTGCTGTTTGTAA
CTTGGCTTCCATTGCCTTGCCATCAT
TTGTTGAAAATGATGAAAAAAGTACTTGGTACAACTTTGACAAATTACATCAG
GTCACTAAGGTTGTCACCCGTAACTTG
AACAGAGTTATTGACCGTAACCATTACCCAGTCCCAGAAGCTGAAAGATCAAA
CATGAGACACAGACCAATTGCTTTGGG
TGTTCAAGGTTTGGCTGATGCCTTTATGGAATTGAGATTACCATTTGACTCTCA
AGAAGCTAGAGAATTGAACATTCAAA
TTTTTGAGACTATCTACCATGCTGCTGTTGAAGCTTCAATTGAATTGGCTAAAG
AAGAAGGTGCCTACGAAACCTATCCA
GGTTCTCCAGCCTCTCAAGGTTTATTACAATTTGATTTGTGGAACAGAAAACCA
ACTGAATTATGGGATTGGGATACATT
AAAACAAGATTTGGCCAAACATGGTATGAGAAACTCCTTGTTGGTTGCACCAA
TGCCTACTGCTTCCACATCACAAATTT
TGGGTAACAATGAATGTTTTGAACCATACACTTCTAACATTTACTCTAGAAGAG
TATTAGCTGGAGAATTCCAAATTGTC
AATCCATATTTATTGAAGGACTTGGTTGATTTGGGTGTCTGGAACGACGCTATG
AAAAGTAGTATTATTGCTAACAATGG
TTCTATCCAAGCCTTACCAAACATCCCTGATGAAATCAAGGCATTGTACAAAA
CTGTCTGGGAAATCTCACAAAAACATA
TTATCGACATGGCTGCTGATAGAGCAGCATTTATTGATCAATCTCAATCATTAA
ACATTCACATCAAAGATCCAACAATG
GGTAAATTAACCAGTATGCACTTCTACGGTTGGAAGAAAGGTTTAAAGACTGG
TATGTACTACTTAAGAACACAAGCTGC
CAGTGCTGCTATTCAATTTACCATTGATCAAAAGATTGCTGAGACTGCCGGTCA
TACGGTTGCAAACTTGGACAAATTAA

Fig 53 (cont)

ACATTAAGAAATATGTTAACAAAGGAAGAGTTGAGAGTGAGAATACCAGTGAT
GCTCCATACAAGTCACCATCAACCGAA
CCAACCTCATTAGAAAGTTCAGTTGCTGATTTGAAAATAAAAGATGAAGGTGA
AAAGCCAGCTGAAGACAAAACCATTGA
AGAACTCGAAAATGACATTTATAGTGCCAAAGTTATCGCATGTGCTATTGATA
ATCCAGAATCTTGTACAATGTGTTCTG
GT

Fig 53 (cont)

>RPL16 (485cL) 759bp in-house: 1-759

GGAGGTNTCNNTCTCTGATTCTTCTCCCCTGCTCCACNCAAGGGCCAACCAACA
ATGAGTCAAGTCGCTCCAAAGTGGTA
CCAATCAGAAAGACGTTCCAGCTNCAAAACAAACCAGAAAAGACTGCTCGTCC
ACAAAAATTACGTGCCTCTTTAGTCCC
AGGTACCGTTTTAATTTTATTGGCCGGTAGATTCAGAGGTAAAAGAGTTGTTTA
CTTGAAGAACTTGGAAGACAACACCT
TATTGGTTTCTGGTCCATTCAAAGTCAATGGTGTTCCATTGAGAAGAGTTAACG
CTAGATACGTTATCGCCACCTCCACC
AAAGTCAACGTTTCTGGTGTTGATGTTTCTAAATTCAACGTCGAATACTTTGCT
AGAGAAAAATCTTCTAAATCTAAAAA
ATCCGAAGCTGAATTCTTCAATGAATCTCAACCAAAGAAAGAAATCAAAGCTG
AAAGAGTTGCTGACCAAAAATCTGTCG
ATGCTGCTTTATTAAGTGAAATCAAAAAGACCCCATTATTGAAACAATACTTG
GCCGCTTCATTCTCTTTGAAGAACGGT
GACAGACCACACTTGTTAAAATTTTAATTTAGGTGAAATTAATATTTTGCAAAC
ATGTTCATGATAAATAACAATGNGGG
CTTTTTAAAGCAATGGGATGGGGATATGGTTAAGAGGGATGGCTTTATATTTTG
AGTTTTTATATATGGGGACCTTTGGT
TTAATAAATGGAANGNTATTGGGCTTCAAAATGAACTTN

Fig 54

>RPS21 (328c3) 391bp in-house: 1-391

AACATTAAAGCAAGATGGAAAACGATAAAGGTCAATTAGTTGAATTATACGTC
CCAAGAAAATGTTCTGCTACCAACAGA
ATCATTAAAGCCAAAGATCACGCTTCTGTTCAAATCTCAATTGCTAAAGTTGAT
GAAGACGGTAGAGCTATTGCTGGTGA
AAACATCACTTACGCTTTAAGTGGTTACGTTAGAGGTAGAGGTGAAGCTGATG
ACTCATTAAACAGATTGGCTCAACAAG
ACGGTTTATTGAAGAACGTCTGGTCTTACTCTCGTTAAGAGAATAGAAGAATA
GACAAAATTGATAATTGGGTATTTTAA
GAAATTACTTTTTTTATATTGCAAATTAATTTTAATCTTTCTTCTGTGTATATTTA
ATGNCTTAACATAAT

Fig 55

>RVS167 (67g1) part1 733bp in-house: 145-733 public: 1-144

TCTACTTCTGCTTGAGCACTGACCATTTTTTCTTCATCTTTAACAGTTCTTTCTTT
CTTCAGTTCATATTTAGAAAAATT
TCTCNTATGACGATCCAAATCCAATTGTTTATGGTCTCTTTTCACTGACATTTTC
CTTATAGCTTGTATAATCTTCAATA
ATTCTTGTGCTGGTTCAACAATTCTTTTTTCAATCAATTCCAAATCGGGTTTTAA
GGTATCTTTGAGATCTTTAACCACT
GCTTGGTACAGTTCCGATGCTTCAATACCTTGTGGGTTATCTTCTGGTACCGTA
GCACTGGGGTCCGATAATCTACCACT
GATTGGTTTATAAATCTCAGCCACGGCTTTGGCAAAATCAATTTGTTCATCTAA
CATCCCATTGACAGCATTGAAATATT
TCTTGGATTCTTCACTCAACTTTTTTGTTTCCRTTTCGATTTCTTTGAATCTTCTT
TCAGCATCGAGATAAACAGCATCT
TGGGTGATTTCTCCCATGTTGAATTTCTGACGCATTGTCTGTGGGGCCCTAAGG
ACACCCTTTTTGAATCCTTTAAATGA
CATAAGATTGTAATAATTGAAAAAATAAAAGGAGAAGGAAGAAGGAAGGGAG
ATAGTATATGAAAAGGAAGGGGCGGAGG
GAATTAATTGTAGGAAGAAGTGGCATTGCTTTTTGTCGAAAGCATTTTTTGAGC
GTGCGAGAAATTTAATCCAAAAAAAT
GTGTGGTGAAAGG

Fig 56(a)

>RVS167 (67g1) part2 1079bp public: 1-523/668-1079 PathoSeq: 524-667

AGTGGTAGATTATCGGACCCCAGTGCTACGGTACCAGAAGATACCCTACAAGG
TATTGAAGCATCGGAACTGTACCAAGC
AGTGGTTANAGATCCTATAGATACTTTAAACCCCGATTTGGAATTGATTGAAAA
AAGAATTGTTGAACCAGCACAAGAAT
TATTGAAGATTATACAAGCTATAAGGAAAATGTCAGTGAAAAGAGACCATAAA
CAATTGGATTTGGATCGTCATAAGAGA
AATTTTTCTAAATATGAACTGAAGAAAGAAAGAACTGTTAAAGATGAAGAAAA
AATGTTCAGTGCTCAAGCAGAAGTAGA
AATTGCTCAACAAGAGTACGATTATTATAATGATTTGTTAAAGAATGAATTGCC
AGTTTTGTATCAAATGCAAAGTGATT
TTATCAAACCATTGTATGTATCATTCTATTACATGCAGTTGAATATTTTCTACAC
ATTATACACTAGAATGGAAGAGTTG
AAAATTCCATATTTTATTTTGTCTACTGATATTGTCGATGCNTATACTGCCAAG
AAGGGGAACATTGAGGAACAAACCGA
TTCTATTGGAATCACTCATTTCAAAGTCGGGCATGCCAAATCCAAATTGGAAGC
CACTAAAAGAAGACATGCTGCTATGA
AATAGTCCACCTCCTACNGGTGCCAAGCTCTATGGCATCTACAGGAACTGGTG
GTGAATTACCTGCATACTCCCCAGGAG
GTTACAACCAACCATATGGTGATAGCAAGTATCAACCACCATCTTCTCCAGCA
ACATACCAATCTCCAGTAGTAGCAGCC
ACTGCTCAATCTCCAGCTACTTATCAATCGCCAGTGGCTACTGGACAACCTCCA
TCATATTTACCACAAACTCCAGCCAG
TGCTCCACCACCACAAGTTGGTAGTGGCCTTCCAACATGCACGGCTTTATACGA
TTATACTGCACAAGCCCAGGGTGACT

Fig 56(b)

TGACTTTCCCTGCAGGAGCTGTTATTGAAATTATACAAAGAACCGAAGATGCC
AACGGATGGTGGACTGGTAAATACAAT
GGTCAAACCGGTGTGTTCCCTGGTAATTATGTGCAATTA

Fig 56(5) (cont)

>SAM2 (36) 1155bp in-house: 1-1155

ATGACTACTTCCAAGGAAACTTTCCTTTTCACTTCAGAATCCGTTGGTGAAGGT
CACCCAGATAAGATTTGTGACCAAGT
CTCCGATGCCATTTTAGATGCTTGTTTAGCTGTTGATCCATTGTCAAAAGTTGCT
TGTGAAACTGCTGCCAAAACCGGTA
TGATTATGGTTTTTGGTGAAATTACCACTAAAGCTCAATTGGATTATCAAAAAA
TCATTAGAGACACCATTAAACACATT
GGTTACGACGATTCTGAAAAAGGTTTTGATTACAAGACTTGTAACGTCTTGGTT
GCAATTGAACAACAATCTCCAGATAT
TGCTCAAGGTTTACATTACGAAAAAGCTTTGGAAGAGTTGGGTGCTGGTGATC
AAGGTATTATGTTTGGTTATGCCACCG
ATGAAACCGATGAAAAATTGCCATTGACCATTTTATTGGCCCACAAATTGAAT
GCTGCCTTGGCTTCTGCCAGAAGATCA
GGTTCCTTGCCATGGTTGAGACCAGATACCAAAACCCAAGTCACCATCGAGTA
TGAAAAAGATGGTGGTGCAGTTATCCC
AAAAAGAGTCGACACAATTGTTATTTCCACTCAACATGCCGAAGAAATCACCA
CCGAAAATTTGAGAAAAGAAATTATTG
AACATATCATCAAGCAAGTCATCCCAGAACATTTATTAGACGACAAAACTATC
TACCACATTCAGCCATCAGGCAGATTC
GTCATTGGTGGTCCCCAAGGTGATGCTGGTTTGACTGGTAGAAAGATCATTGTT
GACACCTATGGTGGTTGGGGTGCACA
TGGTGGTGGTGCCTTCTCAGGCAAGGATTTCTCCAAAGTTGATAGGTCTGCTGC
TTATGCCGCTCGGTGGGTTGCTAAGT
CGTTGGTGACCGCCGGATTGGCCAAAAGGGCCTTGGTGCAGTTCTCCTATGCTA
TTGGGGTTGCTGAACCCACCAGCATT
TATATAGACACCTATGGGACATCTAAATTGAGCACCGAAGCCCTTGTAGAAAT
TATCAAGAATAATTTTGACTTACGCCC
TGGCGTAATTGTAAAAGAATTAGATTTGGCTCGTCCTATTTATTTTAAAACCGC
TTCTTACGGACATTTTACTAACCAAG
AAAATTCTTGGGAACAACCAAAAAAATTAAAATTT

Fig 57

>SAP (232c_cp) 619bp

AACCTATAATTTTCAGAAAGAGACTAGATTCTGATAGAAATATAGACGCATCA
CTATATTTTGGAAATATAGATCCACAA
GTTACGGAGTTGTTAATGTATGAGTTGTTCATCCAATTTGGTCCCGTCAAATCA
ATCAATATGCCAAAGGATCGTATATT
GAAAACACACCAGGGGTATGGATTTGTCGAATTTAAAAACTCAGCAGATGCCA
AATATACTATGGAAATACTACGAGGAA
TAAGACTTTATGGAAAAGCATTGAAATTGAAACGAATTGATGCCAAGTCTCAG
TCATCAACAAACAACCCAAATAATCAA

Fig 58

ACAATAGGAACATTTGTACAATCAGATTTGATCAATCCAAATTACATAGATGTT
GGAGCTAAACTATTTATCAACAATCT
TAATCCATTGGTCGATGAATCCTTTTTAATGGATACGTTTAGTAAGTTTGGAAC
CCTTATAAGAAACCCAATAATTAGAC
GTGATTCAGAGGGACACTCTTTGGGATACGGATTTCTTACGTACGATGACTTTG
AAAGTAGTGATTTATGCATACAAAAA
ATGAACAACACGATTTTGATGAATACCAAAATTGCTATCAGTTATGCATTCAA
NGGATC

Fig 58 (cont)

>SHA3 (83c3) 1376bp in-house: 375-1376 PathoSeq: 1-374

TGNCCTGGAAATCCCCCATTACCATTTTAAAGGTACCACCACCCCCCCAAANCT
TNGCGACTATCCATCCAGGTATTANC
CCTTGGAGGATTNGCCCATAATAATATGGATGGATCATTTGGAGCAAGGAGAT
TTGTCCACTAATATCATGGATAGACAA
ATATCCACCAANAATAGTCATAGAAAAGTTCCAAGAACAGATTTTGAANCCCA
ATTATTAATGAAGAATGCCATGTTACA
ATTGATAGAAGCCATTGAATATTGTCACGAAAATAATATTTACCATTGTGATTT
AAAACCAGAAAACATTATGGTTAGAT
ATAATCCATACTATGTTCGTCCAACTATCAATAACAATAATAACAATGGAGAA
GATGATTTATGCTATGCCAACAGTATT
ATTGACTATAATGAATTACACCTCGTGTTGATTGATTTTGGTTTAGCTATGGAC
TCTGCTACCATTTGTTGTAATTCATG
TCGTGGATCGTCATTTTACATGGCACCAGAAAGAACCACCAATTATAACACCC
ATCGTTTAATCAACCAATTAATTGATA
TGAATCAATATGAGTCAATTGAAATCAATGGGACAACAGTGACAAAATCAAAC
TGTAAATATTTACCTACATTGGCTGGG
GATATTTGGTCATTGGGAGTATTGTTCATTAATATCACTTGTTCAAGAAACCCA
TGGCCCATTGCATCATTTGATAATAA
TCAAAATAATGAAGTGTTTAAGAATTATATGTTGAATAATAACAAGGCTGTTTT
GAGCAAAATCTTACCCATTTCCTCAC
AATTTAATCGCTTATTAGATAGAATTTTCAAATTGAATCCTAATGATAGAATAG
ATTTACCAACTTTATACAAAGAAGTT
ATTCGTTGTGATTTCTTCAAAGATGATCATTACTACTATGCCCAACATCAACAT
CATCACAATCACAATCAAATCAATAA
TGCTTACAATCACTATCAGAAACAACCTAATCAAGCAAGACCTACTGCAAACC
AACAATTGTATACACCACCGGAAACCA
CCACTTATAATTCATACGCTAGTGATATGGAAGAAGATGAAATTAGTGATGAT
GAGTTTTATTCTGATGAAGAAGATGAA
GATATTGAAGACTATGAAGAGGAAGAGGAAGAGTATTTTGGTAATGAGCAAC
AACAACAACAGCAAGTCACAACAGTGAA
TGGTAATTTTGGTCAAGTTAAAGGTACCTGTTATTACGATACCAAAACCAAAA
CAACTACATATATAAAACCACCAGCTG
CATATACTTTAGAGACGCCTAGTCAAAGTGTTGAATACTGTTAAGTTGTACACA
TAAATAATTAATGACAATTAATAATA
ACGATTAATAATATAG

Fig 59

>TPI1 (233c_cp2) 636bp

AACCAATTTTAGAAACAATGGCTCGTCAATTTTTCGTAGGTGGTAACTTCAAAG
CTAACGGTACCAAACAACAAATCACT
TCAATCATCGACAACTTGAACAAGGCTGATTTACCAAAGGATGTCGAAGTTGT
CATTTGTCCACCCGCCCTTTACCTTGG
TTTAGCTGTTGAGCAAAACAAACAACCAACTGTTGCCATTGGTGCTCAAAATG
TTTTTGACAAGTCATGTGGTGCTTTCA
CTGGTGAAACCTGTGCTTCTCAAATCTTGGATGTTGGTGCCAGCTGGACTTTAA
CTGGTCACAGTGAAAGAAGAACCATT
ATNAAAGAATCCGATGAATTCATTGCTGAAAAAACCAAGTTTGCCTTGGACAC
TGGTGTCAAAGTTATTNTATGTATTGG
TGAAACCTTAGAGGAAAGAAAAGGTNGTGTCACTTTGGATGTTTGNGCCAGAC
AATTGGGATGCTGGTTCCAAGATTGNN
TTGATTGGTCAAACATTGNTGGNNCTTACGAACCTGTTTTGGNCAATTGGGTCT
GGTTTANCCCGNTNCCCCANAAGATG
CTGAAGAAACCTACAAGGTNTTAGACTCATTTGGNCAAGANCATTNGTGNCNA
ACAACTGAAAAACCNGANTNTNG

Fig 60

>135g 859bp in-house:1-859
CGTGCATAATTATCTTAAAACCGTAGATAAGCAAAAATTTATCTTATGAAATGTTCAGCGATAAAGAAAGAAAGAATCAG
GTACCACGAGGAGTGTTTTTGAGAAAAACAACTCGTAAATTAATGAATCTAGTTTCTCTATACTTGAATAATTTTTGAGT
TTTCTGGAAAAGACACCTGTTCCAGTTTCAAATTAAACAAGAATGTGAAAAGAATAAAATTTGATTTATTCTAGCCTGTT
AATAATCCAGGAAAACTCAATTTTCGTAATTGGCAACTTGTCCGAGTGGTTAAGGAGAAAGATTAGAAATCTTTTGGGCT
TTGCCCGCGCAGGTTCGAGTCCTGCAGTTGTCGTTATTTTTTTTGGTTTACTCTCTATTTTAAAATTTAAAACTAATCAA
CTGAAACTGGAGTACCTGCCATGATATGAGTAAATACTTTTTTGATATTAAAAATCTATATAAAACTCCCTATTTATTTT
TTAATTTAAACCCAGATATTGTCCCAATAATAGTTTTTTGTTTGAACTTATTGCTTTGTATGAACCTTGTTAGTTTAATC
TTTCCAATTTCATACTCTCTTAGTTGGCCACATCAGTGGCTCATTGAATAATTCTGATCTTGAAGTGTACCAGATGTATT
CTGACAAAACTGCACACGGACCCAGTCAATAGCATTATAGATATTTTGATTTAAAGTTCACCGAATATATCGAATATCTT
TATTGGCCATCTCATCTCATCTTCTTGCAATAAATTCTTAAACGCTACTTTTTCTCAAACCTTATTATCCCTCTAGATAC
TCTTCCAAATCTTCAGGTTCAAATATCACTTTAACCATCAATGAACAACTAGGGCAAAC

Fig 61

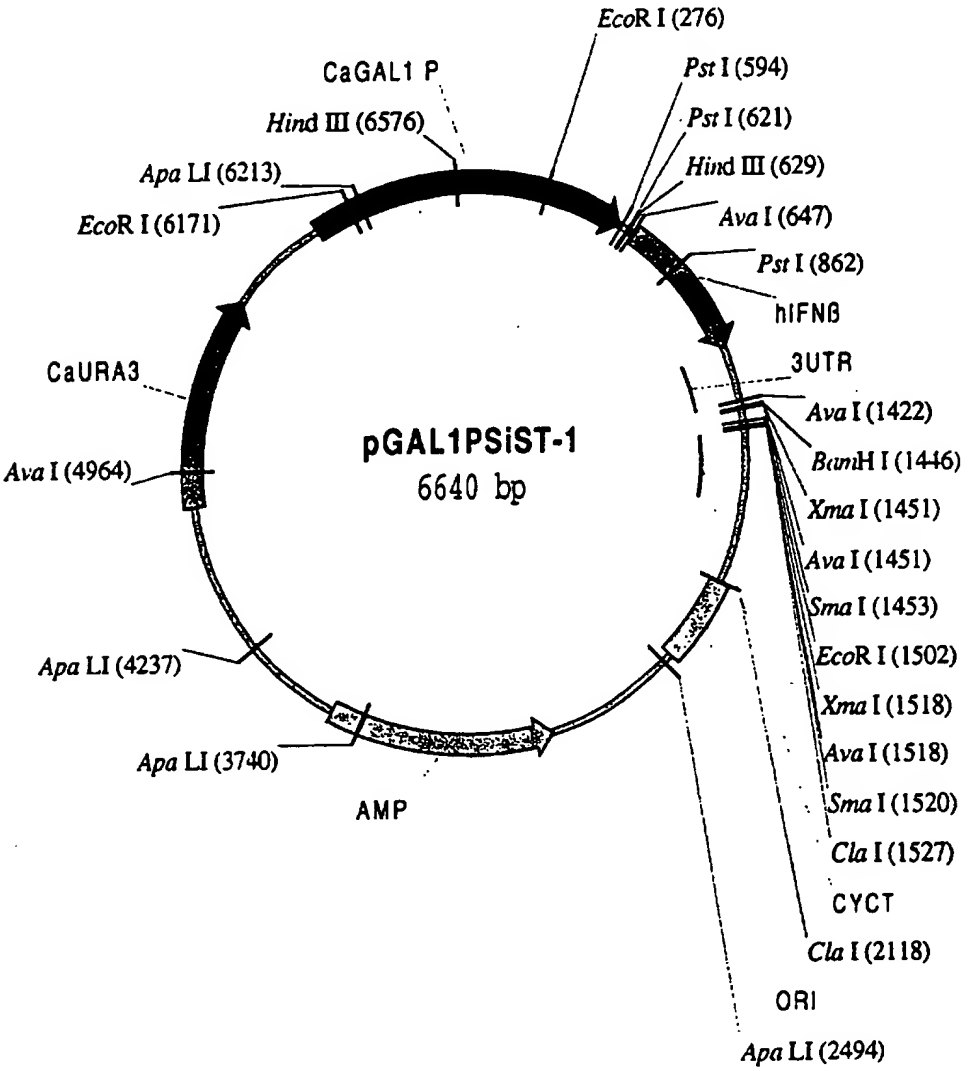

pGAL1PSiST-1
6640 bp
EcoR I (276)
Pst I (594)
Pst I (621)
Hind III (629)
Ava I (647)
Pst I (862)
hIFNß
3UTR
Ava I (1422)
BamH I (1446)
Xma I (1451)
Ava I (1451)
Sma I (1453)
EcoR I (1502)
Xma I (1518)
Ava I (1518)
Sma I (1520)
Cla I (1527)
CYCT
Cla I (2118)
ORI
Apa LI (2494)
AMP
Apa LI (3740)
Apa LI (4237)
Ava I (4964)
CaURA3
Hind III (6576)
CaGAL1 P
Apa LI (6213)
EcoR I (6171)

Fig 62

```
   1  TTCCATCGGG GAAAGTGGGG GGGAAAAAAT TTTAAGCAGT TCACAAAACC
      AAGGTAGCCC CTTTCACCCC CCCTTTTTTA AAATTCGTCA AGTGTTTTGG
 .....
  51  TTCCAAAAAA TATATGGACA AAGATGATTG TATTTTCCCG ACACCAAAAT
      AAGGTTTTTT ATATACCTGT TTCTACTAAC ATAAAAGGGC TGTGGTTTTA
 .....
 101  CATAATTAAT TATGAGAAAG TTAAATGTAA CGTTACAATT TATGTTTATT
      GTATTAATTA ATACTCTTTC AATTTACATT GCAATGTTAA ATACAAATAA
 .....
 151  TGAAGGTGAA AAGCGATTTA TGATTTTTCC GAAATGAAAA TTTTTTTTAG
      ACTTCCACTT TTCGCTAAAT ACTAAAAAGG CTTTACTTTT AAAAAAAATC
 .....
 201  GTTTATTTTT TTTGTCGGGC AAAGAAAAAC TGAACAAGGA TTATTAAAAT
      CAAATAAAAA AAACAGCCCG TTTCTTTTTG ACTTGTTCCT AATAATTTTA
 .....
                                EcoRI
                               ------
 251  TTTTGGTGTT TGTTTGTGTC TGGAGAATTC ATTCCTCTCT CATCTTCACA
      AAAACCACAA ACAAACACAG ACCTCTTAAG TAAGGAGAGA GTAGAAGTGT
 .....
 301  CAATGTTTAG ACATCTGACA CGATTCATGA TAGTTCGGTT TCCGGGGTTG
      GTTACAAATC TGTAGACTGT GCTAAGTACT ATCAAGCCAA AGGCCCCAAC
 .....
 351  GTGTTTAGTT TTCGTTTTTC TTTTTTTTTG GAAAGAATGT TTTAGCTCAT
      CACAAATCAA AAGCAAAAAG AAAAAAAAAC CTTTCTTACA AAATCGAGTA
 .....
 401  TGGTTTTCTT TCTTCATTCA ATAGTTTTGA AAGAATTTGC CCACTTGTTA
      ACCAAAAGAA AGAAGTAAGT TATCAAAACT TTCTTAAACG GGTGAACAAT
 .....
 451  TTACAATCAT ATAAAATTAA ACTTTGATAT AAAATAGAGT TTGAAAGTTT
      AATGTTAGTA TATTTTAATT TGAAACTATA TTTTATCTCA AACTTTCAAA
 .....
 501  CCCAGATCCT TTTTGATTTC TTTGTAAATT TTTTTTTCTC CCACATATAC
      GGGTCTAGGA AAAACTAAAG AAACATTTAA AAAAAAAGAG GGTGTATATG
 .....
                                          PstI
                                         -------
 551  ACACATACAA ACCGATTTTT ATAAGAAAGA GTTATACCCT GCAGCTCGAC
      TGTGTATGTT TGGCTAAAAA TATTCTTTCT CAATATGGGA CGTCGAGCTG
 .....
                     PstI         HindIII              AvaI
                    -------      -------              -----
 601  CTCGACTGTT TAAACCTGCA GGCATGCAAG CTTGGCCAAA AAGGCCTCGA
      GAGCTGACAA ATTTGGACGT CCGTACGTTC GAACCGGTTT TTCCGGAGCT
 .....
      AvaI
      -
 651  GGAACATGAC CAACAAGTGT CTCCTCCAAA TTGCTCTCCT GTTGTGCTTC
      CCTTGTACTG GTTGTTCACA GAGGAGGTTT AACGAGAGGA CAACACGAAG
 .....
 701  TCCACTACAG CTCTTTCCAT GAGCTACAAC TTGCTTGGAT TCCTACAAAG
      AGGTGATGTC GAGAAAGGTA CTCGATGTTG AACGAACCTA AGGATGTTTC
 .....
 751  AAGCAGCAAT TTTCAGTGTC AGAAGCTCCT GTGGCAATTG AATGGGAGGC
      TTCGTCGTTA AAAGTCACAG TCTTCGAGGA CACCGTTAAC TTACCCTCCG
 .....
 801  TTGAATACTG CCTCAAGGAC AGGATGAACT TTGACATCCC TGAGGAGATT
      AACTTATGAC GGAGTTCCTG TCCTACTTGA AACTGTAGGG ACTCCTCTAA
 .....
```

Fig 63

```
            PstI
            -------
 851  AAGCAGCTGC AGCAGTTCCA GAAGGAGGAC GCCGCATTGA CCATCTATGA
      TTCGTCGACG TCGTCAAGGT CTTCCTCCTG CGGCGTAACT GGTAGATACT
 ...........................................................................
 901  GATGCTCCAG AACATCTTTG CTATTTTCAG ACAAGATTCA TCTAGCACTG
      CTACGAGGTC TTGTAGAAAC GATAAAAGTC TGTTCTAAGT AGATCGTGAC
 ...........................................................................
 951  GCTGGAATGA GACTATTGTT GAGAACCTCC TGGCTAATGT CTATCATCAG
      CGACCTTACT CTGATAACAA CTCTTGGAGG ACCGATTACA GATAGTAGTC
 ...........................................................................
1001  ATAAACCATC TGAAGACAGT CCTGGAAGAA AAACTGGAGA AAGAAGATTT
      TATTTGGTAG ACTTCTGTCA GGACCTTCTT TTTGACCTCT TTCTTCTAAA
 ...........................................................................
1051  CACCAGGGGA AAACTCATGA GCAGTCTGCA CCTGAAAAGA TATTATGGGA
      GTGGTCCCCT TTTGAGTACT CGTCAGACGT GGACTTTTCT ATAATACCCT
 ...........................................................................
1101  GGATTCTGCA TTACCTGAAG GCCAAGGAGT ACAGTCACTG TGCCTGGACC
      CCTAAGACGT AATGGACTTC CGGTTCCTCA TGTCAGTGAC ACGGACCTGG
 ...........................................................................
1151  ATAGTCAGAG TGGAAATCCT AAGGAACTTT TACTTCATTA ACAGACTTAC
      TATCAGTCTC ACCTTTAGGA TTCCTTGAAA ATGAAGTAAT TGTCTGAATG
 ...........................................................................
1201  AGGTTACCTC CGAAACTGAA GATCTCCTAG CCTGTGCCTC TGGGACTGGA
      TCCAATGGAG GCTTTGACTT CTAGAGGATC GGACACGGAG ACCCTGACCT
 ...........................................................................
1251  CAATTGCTTC AAGCATTCTT CAACCAGCAG ATGCTGTTTA AGTGACTGAT
      GTTAACGAAG TTCGTAAGAA GTTGGTCGTC TACGACAAAT TCACTGACTA
 ...........................................................................
1301  GGCTAATGTA CTGCATATGA AAGGACACTA GAAGATTTTG AAATTTTTAT
      CCGATTACAT GACGTATACT TTCCTGTGAT CTTCTAAAAC TTTAAAAATA
 ...........................................................................
1351  TAAATTATGA GTTATTTTTA TTTATTTAAA TTTTATTTTG GAAAATAAAT
      ATTTAATACT CAATAAAAAT AAATAAATTT AAAATAAAAC CTTTTATTTA
 ...........................................................................
                                                      XmaI
                                                       ~
                                                      SmaI
                                                       ~
                                                     BamHI
                                                     ------
                                 AvaI                 AvaI
                                 ------                ~
1401  TATTTTTGGT GCAAAAGTCC CTCGAGGCCT AGCGGCCGCC TAGAGGATCC
      ATAAAAACCA CGTTTTCAGG GAGCTCCGGA TCGCCGGCGG ATCTCCTAGG
 ...........................................................................
      XmaI
      -----
      SmaI
      -----
      AvaI
      -----
1451  CCGGGCGCTA GGCGGCCGCT AGGCCTTTTT GGCCAAGCTC GAATTTCGAG
      GGCCCGCGAT CCGCCGGCGA TCCGGAAAAA CCGGTTCGAG CTTAAAGCTC
 ...........................................................................
                       XmaI
                      -------
                       SmaI
                      -------
 EcoRI                 AvaI     ClaI
 ------               -------  ------
1501  GAATTCGAGC TCGGTACCCG GGGGATCGAT CCGTCCCCCT TTTCCTTTGT
      CTTAAGCTCG AGCCATGGGC CCCCTAGCTA GGCAGGGGGA AAAGGAAACA
 ...........................................................................
```

```
1551  CGATATCATG TAATTAGTTA TGTCACGCTT ACATTCACGC CCTCCCCCCA
      GCTATAGTAC ATTAATCAAT ACAGTGCGAA TGTAAGTGCG GGAGGGGGGT
.....

1601  CATCCGCTCT AACCGAAAAG GAAGGAGTTA GACAACCTGA AGTCTAGGTC
      GTAGGCGAGA TTGGCTTTTC CTTCCTCAAT CTGTTGGACT TCAGATCCAG
.....

1651  CCTATTTATT TTTTTATAGT TATGTTAGTA TTAAGAACGT TATTTATATT
      GGATAAATAA AAAAATATCA ATACAATCAT AATTCTTGCA ATAAATATAA
.....

1701  TCAAATTTTT CTTTTTTTTC TGTACAGACG CGTGTACGCA TGTAACATTA
      AGTTTAAAAA GAAAAAAAAG ACATGTCTGC GCACATGCGT ACATTGTAAT
.....

1751  TACTGAAAAC CTTGCTTGAG AAGGTTTTGG GACGCTCGAA GGCTTTAATT
      ATGACTTTTG GAACGAACTC TTCCAAAACC CTGCGAGCTT CCGAAATTAA
.....

1801  TGCAAGCTAG CTTGGCGTAA TCATGGTCAT AGCTGTTTCC TGTGTGAAAT
      ACGTTCGATC GAACCGCATT AGTACCAGTA TCGACAAAGG ACACACTTTA
.....

1851  TGTTATCCGC TCACAATTCC ACACAACATA CGAGCCGGAA GCATAAAGTG
      ACAATAGGCG AGTGTTAAGG TGTGTTGTAT GCTCGGCCTT CGTATTTCAC
.....

1901  TAAAGCCTGG GGTGCCTAAT GAGTGAGCTA ACTCACATTA ATTGCGTTGC
      ATTTCGGACC CCACGGATTA CTCACTCGAT TGAGTGTAAT TAACGCAACG
.....

1951  GCTCACTGCC CGCTTTCCAG TCGGGAAACC TGTCGTGCCA GAGATCTCTG
      CGAGTGACGG GCGAAAGGTC AGCCCTTTGG ACAGCACGGT CTCTAGAGAC
.....

2001  CATTAATGAA TCGGCCAACG CGCGGGGAGA GGCGGTTTGC GTATTGGGCG
      GTAATTACTT AGCCGGTTGC GCGCCCCTCT CCGCCAAACG CATAACCCGC
.....

2051  CTCTTCCGCT TCCTCGCTCA CTGACTCGCT GCGCTCGGTC GTTCGGCTGC
      GAGAAGGCGA AGGAGCGAGT GACTGAGCGA CGCGAGCCAG CAAGCCGACG
.....

                    ClaI
                    -------
2101  GGCGAGCGGT ATCAGATCGA TCTCACTCAA AGGCGGTAAT ACGGTTATCC
      CCGCTCGCCA TAGTCTAGCT AGAGTGAGTT TCCGCCATTA TGCCAATAGG
.....

2151  ACAGAATCAG GGGATAACGC AGGAAAGAAC ATGTGAGCAA AAGGCCAGCA
      TGTCTTAGTC CCCTATTGCG TCCTTTCTTG TACACTCGTT TTCCGGTCGT
.....

2201  AAAGGCCAGG AACCGTAAAA AGGCCGCGTT GCTGGCGTTT TTCCATAGGC
      TTTCCGGTCC TTGGCATTTT TCCGGCGCAA CGACCGCAAA AAGGTATCCG
.....

2251  TCCGCCCCCC TGACGAGCAT CACAAAAATC GACGCTCAAG TCAGAGGTGG
      AGGCGGGGGG ACTGCTCGTA GTGTTTTTAG CTGCGAGTTC AGTCTCCACC
.....

2301  CGAAACCCGA CAGGACTATA AAGATACCAG GCGTTTCCCC CTGGAAGCTC
      GCTTTGGGCT GTCCTGATAT TTCTATGGTC CGCAAAGGGG GACCTTCGAG
.....

2351  CCTCGTGCGC TCTCCTGTTC CGACCCTGCC GCTTACCGGA TACCTGTCCG
      GGAGCACGCG AGAGGACAAG GCTGGGACGG CGAATGGCCT ATGGACAGGC
.....

2401  CCTTTCTCCC TTCGGGAAGC GTGGCGCTTT CTCATAGCTC ACGCTGTAGG
      GGAAAGAGGG AAGCCCTTCG CACCGCGAAA GAGTATCGAG TGCGACATCC
.....

                                                    ApaLI
                                                    ------
2451  TATCTCAGTT CGGTGTAGGT CGTTCGCTCC AAGCTGGGCT GTGTGCACGA
      ATAGAGTCAA GCCACATCCA GCAAGCGAGG TTCGACCCGA CACACGTGCT
.....
```

```
2501  ACCCCCCGTT CAGCCCGACC GCTGCGCCTT ATCCGGTAAC TATCGTCTTG
      TGGGGGGCAA GTCGGGCTGG CGACGCGGAA TAGGCCATTG ATAGCAGAAC
.....
2551  AGTCCAACCC GGTAAGACAC GACTTATCGC CACTGGCAGC AGCCACTGGT
      TCAGGTTGGG CCATTCTGTG CTGAATAGCG GTGACCGTCG TCGGTGACCA
.....
2601  AACAGGATTA GCAGAGCGAG GTATGTAGGC GGTGCTACAG AGTTCTTGAA
      TTGTCCTAAT CGTCTCGCTC CATACATCCG CCACGATGTC TCAAGAACTT
.....
2651  GTGGTGGCCT AACTACGGCT ACACTAGAAG GACAGTATTT GGTATCTGCG
      CACCACCGGA TTGATGCCGA TGTGATCTTC CTGTCATAAA CCATAGACGC
.....
2701  CTCTGCTGAA GCCAGTTACC TTCGGAAAAA GAGTTGGTAG CTCTTGATCC
      GAGACGACTT CGGTCAATGG AAGCCTTTTT CTCAACCATC GAGAACTAGG
.....
2751  GGCAAACAAA CCACCGCTGG TAGCGGTGGT TTTTTTGTTT GCAAGCAGCA
      CCGTTTGTTT GGTGGCGACC ATCGCCACCA AAAAAACAAA CGTTCGTCGT
.....
2801  GATTACGCGC AGAAAAAAAG GATCTCAAGA AGATCCTTTG ATCTTTTCTA
      CTAATGCGCG TCTTTTTTTC CTAGAGTTCT TCTAGGAAAC TAGAAAAGAT
.....
2851  CGGGGTCTGA CGCTCAGTGG AACGAAAACT CACGTTAAGG GATTTTGGTC
      GCCCCAGACT GCGAGTCACC TTGCTTTTGA GTGCAATTCC CTAAAACCAG
.....
2901  ATGAGATTAT CAAAAAGGAT CTTCACCTAG ATCCTTTTAA ATTAAAAATG
      TACTCTAATA GTTTTTCCTA GAAGTGGATC TAGGAAAATT TAATTTTTAC
.....
2951  AAGTTTTAAA TCAATCTAAA GTATATATGA GTAAACTTGG TCTGACAGTT
      TTCAAAATTT AGTTAGATTT CATATATACT CATTTGAACC AGACTGTCAA
.....
3001  ACCAATGCTT AATCAGTGAG GCACCTATCT CAGCGATCTG TCTATTTCGT
      TGGTTACGAA TTAGTCACTC CGTGGATAGA GTCGCTAGAC AGATAAAGCA
.....
3051  TCATCCATAG TTGCCTGACT CCCCGTCGTG TAGATAACTA CGATACGGGA
      AGTAGGTATC AACGGACTGA GGGGCAGCAC ATCTATTGAT GCTATGCCCT
.....
3101  GGGCTTACCA TCTGGCCCCA GTGCTGCAAT GATACCGCGA GACCCACGCT
      CCCGAATGGT AGACCGGGGT CACGACGTTA CTATGGCGCT CTGGGTGCGA
.....
3151  CACCGGCTCC AGATTTATCA GCAATAAACC AGCCAGCCGG AAGGGCCGAG
      GTGGCCGAGG TCTAAATAGT CGTTATTTGG TCGGTCGGCC TTCCCGGCTC
.....
3201  CGCAGAAGTG GTCCTGCAAC TTTATCCGCC TCCATCCAGT CTATTAATTG
      GCGTCTTCAC CAGGACGTTG AAATAGGCGG AGGTAGGTCA GATAATTAAC
.....
3251  TTGCCGGGAA GCTAGAGTAA GTAGTTCGCC AGTTAATAGT TTGCGCAACG
      AACGGCCCTT CGATCTCATT CATCAAGCGG TCAATTATCA AACGCGTTGC
.....
3301  TTGTTGCCAT TGCTACAGGC ATCGTGGTGT CACGCTCGTC GTTTGGTATG
      AACAACGGTA ACGATGTCCG TAGCACCACA GTGCGAGCAG CAAACCATAC
.....
3351  GCTTCATTCA GCTCCGGTTC CCAACGATCA AGGCGAGTTA CATGATCCCC
      CGAAGTAAGT CGAGGCCAAG GGTTGCTAGT TCCGCTCAAT GTACTAGGGG
.....
3401  CATGTTGTGC AAAAAAGCGG TTAGCTCCTT CGGTCCTCCG ATCGTTGTCA
      GTACAACACG TTTTTTCGCC AATCGAGGAA GCCAGGAGGC TAGCAACAGT
.....
3451  GAAGTAAGTT GGCCGCAGTG TTATCACTCA TGGTTATGGC AGCACTGCAT
      CTTCATTCAA CCGGCGTCAC AATAGTGAGT ACCAATACCG TCGTGACGTA
.....
```

```
3501  AATTCTCTTA CTGTCATGCC ATCCGTAAGA TGCTTTTCTG TGACTGGTGA
      TTAAGAGAAT GACAGTACGG TAGGCATTCT ACGAAAAGAC ACTGACCACT
.....
3551  GTACTCAACC AAGTCATTCT GAGAATAGTG TATGCGGCGA CCGAGTTGCT
      CATGAGTTGG TTCAGTAAGA CTCTTATCAC ATACGCCGCT GGCTCAACGA
.....
3601  CTTGCCCGGC GTCAATACGG GATAATACCG CGCCACATAG CAGAACTTTA
      GAACGGGCCG CAGTTATGCC CTATTATGGC GCGGTGTATC GTCTTGAAAT
.....
3651  AAAGTGCTCA TCATTGGAAA ACGTTCTTCG GGGCGAAAAC TCTCAAGGAT
      TTTCACGAGT AGTAACCTTT TGCAAGAAGC CCCGCTTTTG AGAGTTCCTA
.....
                                          ApaLI
                                          -------
3701  CTTACCGCTG TTGAGATCCA GTTCGATGTA ACCCACTCGT GCACCCAACT
      GAATGGCGAC AACTCTAGGT CAAGCTACAT TGGGTGAGCA CGTGGGTTGA
.....
3751  GATCTTCAGC ATCTTTTACT TTCACCAGCG TTTCTGGGTG AGCAAAAACA
      CTAGAAGTCG TAGAAAATGA AAGTGGTCGC AAAGACCCAC TCGTTTTTGT
.....
3801  GGAAGGCAAA ATGCCGCAAA AAAGGGAATA AGGGCGACAC GGAAATGTTG
      CCTTCCGTTT TACGGCGTTT TTTCCCTTAT TCCCGCTGTG CCTTTACAAC
.....
3851  AATACTCATA CTCTTCCTTT TTCAATATTA TTGAAGCATT TATCAGGGTT
      TTATGAGTAT GAGAAGGAAA AAGTTATAAT AACTTCGTAA ATAGTCCCAA
.....
3901  ATTGTCTCAT GAGCGGATAC ATATTTGAAT GTATTTAGAA AAATAAACAA
      TAACAGAGTA CTCGCCTATG TATAAACTTA CATAAATCTT TTTATTTGTT
.....
3951  ATAGGGGTTC CGCGCACATT TCCCCGAAAA GTGCCACCTG ACGTCTAAGA
      TATCCCCAAG GCGCGTGTAA AGGGGCTTTT CACGGTGGAC TGCAGATTCT
.....
4001  AACCATTATT ATCATGACAT TAACCTATAA AAATAGGCGT ATCACGAGGC
      TTGGTAATAA TAGTACTGTA ATTGGATATT TTTATCCGCA TAGTGCTCCG
.....
4051  CCTTTCGTCT CGCGCGTTTC GGTGATGACG GTGAAAACCT CTGACACATG
      GGAAAGCAGA GCGCGCAAAG CCACTACTGC CACTTTTGGA GACTGTGTAC
.....
4101  CAGCTCCCGG AGACGGTCAC AGCTTGTCTG TAAGCGGATG CCGGGAGCAG
      GTCGAGGGCC TCTGCCAGTG TCGAACAGAC ATTCGCCTAC GGCCCTCGTC
.....
4151  ACAAGCCCGT CAGGGCGCGT CAGCGGGTGT TGGCGGGTGT CGGGGCTGGC
      TGTTCGGGCA GTCCCGCGCA GTCGCCCACA ACCGCCCACA GCCCCGACCG
.....
                                          ApaLI
                                          -------
4201  TTAACTATGC GGCATCAGAG CAGATTGTAC TGAGAGTGCA CCATATCGAC
      AATTGATACG CCGTAGTCTC GTCTAACATG ACTCTCACGT GGTATAGCTG
.....
4251  GCTCTCCCTT ATGCGACTCC TGCATTAGGA AGCAGCCCAG TAGTAGGTTG
      CGAGAGGGAA TACGCTGAGG ACGTAATCCT TCGTCGGGTC ATCATCCAAC
.....
4301  AGGCCGTTGA GCACCGCCGC CGCAAGGAAT GGTGCATGCA AGGAGATGGC
      TCCGGCAACT CGTGGCGGCG GCGTTCCTTA CCACGTACGT TCCTCTACCG
.....
4351  GCCCAACAGT CCCCCGGCCA CGGGGCCTGC CACCATACCC ACGCCGAAAC
      CGGGTTGTCA GGGGGCCGGT GCCCCGGACG GTGGTATGGG TGCGGCTTTG
.....
4401  AAGCACTAAT AGGAATTGAT TTGGATGGTA TAAACGGAAA CAAAAAAAAG
      TTCGTGATTA TCCTTAACTA AACCTACCAT ATTTGCCTTT GTTTTTTTTC
.....
```

```
4451  AGCTGGTACT ACTTTCTTTA AAATTATTTT ATTATTTGAT TTTATTTAAT
      TCGACCATGA TGAAAGAAAT TTTAATAAAA TAATAAACTA AAATAAATTA
.....
4501  AGTATATATT ATATTTTGAA CGTAGATTAT TTTGTTGAAA GTTGCTGTAG
      TCATATATAA TATAAAACTT GCATCTAATA AAACAACTTT CAACGACATC
.....
4551  TGCCATTGAT TCGTAACACT AATTCTGTAT TAGTCATTCC TCTTGTTTGA
      ACGGTAACTA AGCATTGTGA TTAAGACATA ATCAGTAAGG AGAACAAACT
.....
4601  TAGTATCCAA AAAAACGGCT ATTTTTTTGC AATCTTATTT CCTGCATATT
      ATCATAGGTT TTTTTGCCGA TAAAAAAACG TTAGAATAAA GGACGTATAA
.....
4651  ATACAGATAA CATAATGAAA GAAAAAATCT TTTTTTTTGT TCTTCAATGA
      TATGTCTATT GTATTACTTT CTTTTTTAGA AAAAAAAACA AGAAGTTACT
.....
4701  TGATTTCAAC CATTCTTTTA AACATTGATC AATTCCTGAG CAACAACCCC
      ACTAAAGTTG GTAAGAAAAT TTGTAACTAG TTAAGGACTC GTTGTTGGGG
.....
4751  ATACACACTG GTTTATATAC CGCCCCTTTT ACAGTTGAAG AAAGAAATAG
      TATGTGTGAC CAAATATATG GCGGGGAAAA TGTCAACTTC TTTCTTTATC
.....
4801  AAATAGAAAT AGCAAACAAA AGATATGACA GTCAACACTA AGACCTATAG
      TTTATCTTTA TCGTTTGTTT TCTATACTGT CAGTTGTGAT TCTGGATATC
.....
4851  TGAGAGAGCA GAAACTCATG CCTCACCAGT AGCACAGCGA TTATTTCGAT
      ACTCTCTCGT CTTTGAGTAC GGAGTGGTCA TCGTGTCGCT AATAAAGCTA
.....
4901  TAATGGAACT GAAGAAAACC AATTTATGTG CATCAATTGA CGTTGATACC
      ATTACCTTGA CTTCTTTTGG TTAAATACAC GTAGTTAACT GCAACTATGG
.....
                  AvaI
                  ------
4951  ACTAAGGAGT TCCTCGAGTT AATTGATAAA TTAGGTCCTT ATGTATGCTT
      TGATTCCTCA AGGAGCTCAA TTAACTATTT AATCCAGGAA TACATACGAA
.....
5001  AATCAAGACT CATATTGATA TAATCAATGA TTTTTCCTAT GAATCCACTA
      TTAGTTCTGA GTATAACTAT ATTAGTTACT AAAAAGGATA CTTAGGTGAT
.....
5051  TTGAACCATT ATTAGAACTT TCACGTAAAC ATCAATTTAT GATTTTTGAA
      AACTTGGTAA TAATCTTGAA AGTGCATTTG TAGTTAAATA CTAAAAACTT
.....
5101  GATAGAAAAT TTGCTGATAT TGGTAATACC GTAAAGAAAC AATATATTGG
      CTATCTTTTA AACGACTATA ACCATTATGG CATTTCTTTG TTATATAACC
.....
5151  TGGAGTTTAT AAAATTAGTA GTTGGGCAGA TATTACCAAT GCTCATGGTG
      ACCTCAAATA TTTTAATCAT CAACCCGTCT ATAATGGTTA CGAGTACCAC
.....
5201  TCACTGGGAA TGGAGTGGTT GAAGGATTAA AACAGGGAGC TAAAGAAACC
      AGTGACCCTT ACCTCACCAA CTTCCTAATT TTGTCCCTCG ATTTCTTTGG
.....
5251  ACCACCAACC AAGAGCCAAG AGGGTTATTG ATGTTAGCTG AATTATCATC
      TGGTGGTTGG TTCTCGGTTC TCCCAATAAC TACAATCGAC TTAATAGTAG
.....
5301  AGTGGGATCA TTAGCATATG GAGAATATTC TCAAAAAACT GTTGAAATTG
      TCACCCTAGT AATCGTATAC CTCTTATAAG AGTTTTTTGA CAACTTTAAC
.....
5351  CTAAATCCGA TAAGGAATTT GTTATTGGAT TTATTGCCCA ACGTGATATG
      GATTTAGGCT ATTCCTTAAA CAATAACCTA AATAACGGGT TGCACTATAC
.....
```

```
5401  GGTGGCCAAG AAGAAGGATT TGATTGGCTT ATTATGACAC CTGGAGTTGG
      CCACCGGTTC TTCTTCCTAA ACTAACCGAA TAATACTGTG GACCTCAACC
 .....
5451  ATTAGATGAT AAAGGTGATG GATTAGGACA ACAATATAGA ACTGTTGATG
      TAATCTACTA TTTCCACTAC CTAATCCTGT TGTTATATCT TGACAACTAC
 .....
5501  AAGTTGTTAG CACTGGAACT GATATTATCA TTGTTGGTAG AGGATTGTTT
      TTCAACAATC GTGACCTTGA CTATAATAGT AACAACCATC TCCTAACAAA
 .....
5551  GGTAAAGGAA GAGATCCAGA TATTGAAGGT AAAAGGTATA GAAATGCTGG
      CCATTTCCTT CTCTAGGTCT ATAACTTCCA TTTTCCATAT CTTTACGACC
 .....
5601  TTGGAATGCT TATTTGAAAA AGACTGGCCA ATTATAAATG TGAAGGGGGA
      AACCTTACGA ATAAACTTTT TCTGACCGGT TAATATTTAC ACTTCCCCCT
 .....
5651  GATTTTCACT TTATTAGATT TGTATATATG TAGAATAAAT AAATAAATAA
      CTAAAAGTGA AATAATCTAA ACATATATAC ATCTTATTTA TTTATTTATT
 .....
5701  GTTAAATAAA TAATTAAATA AGGGTGGTAA TTATTACTAT TTACAATCAA
      CAATTTATTT ATTAATTTAT TCCCACCATT AATAATGATA AATGTTAGTT
 .....
5751  AGGTGGTCCT TCTAGCTGTA ATCCGGGCAG CGCAACGGAA CATTCATCAG
      TCCACCAGGA AGATCGACAT TAGGCCCGTC GCGTTGCCTT GTAAGTAGTC
 .....
5801  TGTAAAAATG GAATCAATAA AGCCCTGCGC TCATGAGCCC GAAGTGGCGA
      ACATTTTTAC CTTAGTTATT TCGGGACGCG AGTACTCGGG CTTCACCGCT
 .....
5851  GCCCGATCTT CCCCATCGGT GATGTCGGCG ATATAGGCGC CAGCAACCGC
      CGGGCTAGAA GGGGTAGCCA CTACAGCCGC TATATCCGCG GTCGTTGGCG
 .....
5901  ACCTGTGGCG CCGCAGCGCG CAGGGTCAGC CTGAATACGC GTTTAATGAC
      TGGACACCGC GGCGTCGCGC GTCCCAGTCG GACTTATGCG CAAATTACTG
 .....
5951  CAGCACAGTC GTGATGGCAA GGTCAGAATA GCCCAAGTCG GCCGAGGGGC
      GTCGTGTCAG CACTACCGTT CCAGTCTTAT CGGGTTCAGC CGGCTCCCCG
 .....
6001  CTGTACAGTG AGGGAAGATC TGATATTGAC GAAGAGGAAC CAATGTAACG
      GACATGTCAC TCCCTTCTAG ACTATAACTG CTTCTCCTTG GTTACATTGC
 .....
6051  TTACACTGAA GAAAACACAC AATAAACGGG AAGAAACGGT GTAAAAGTGT
      AATGTGACTT CTTTTGTGTG TTATTTGCCC TTCTTTGCCA CATTTTCACA
 .....
6101  GAAAATAATT TTTGAATATC ATTTCCCTTG GTTTAATTCC AAACGAAACG
      CTTTTATTAA AAACTTATAG TAAAGGGAAC CAAATTAAGG TTTGCTTTGC
 .....
                          EcoRI
                          -------
6151  TGTTTTTTTT AGAGAATGGG AATTCTTATT GGATGTCTAG ATTGTTTGTT
      ACAAAAAAAA TCTCTTACCC TTAAGAATAA CCTACAGATC TAACAAACAA
 .....
                ApaLI
                ------
6201  TACTCCAGAC TGTGCACAAA AACGTTTGGA TGGATGATCA GAAGATATTT
      ATGAGGTCTG ACACGTGTTT TTGCAAACCT ACCTACTAGT CTTCTATAAA
 .....
6251  TTAGGCTTAG CTCTAAATAT AAGAAATGAT GCTTGAAAAA CCAGACAGAA
      AATCCGAATC GAGATTTATA TTCTTTACTA CGAACTTTTT GGTCTGTCTT
 .....
6301  ATTGAGTTTC AAAAATTGGT AATGTGAGGT ATTAGTCAAC TAACCAAATA
      TAACTCAAAG TTTTTAACCA TTACACTCCA TAATCAGTTG ATTGGTTTAT
 .....
```

```
6351  ACAATGCAAA CCGGTTGATA CATTTCATTT TGAAAATAAT GAAACTGGAA
      TGTTACGTTT GGCCAACTAT GTAAAGTAAA ACTTTTATTA CTTTGACCTT
.....

6401  TTGGATGACC AGCACACAAA CACATAAAGT AATTATGGGA ATTAGAAGCG
      AACCTACTGG TCGTGTGTTT GTGTATTTCA TTAATACCCT TAATCTTCGC
.....

6451  AACATAGAGG AGTACTTGGC CACGAACAGA ATACAAGTGG GAACACTATT
      TTGTATCTCC TCATGAACCG GTGCTTGTCT TATGTTCACC CTTGTGATAA
.....

6501  TTCTCCATTG TTTTAGTTCT GTTTTTTTGT CAGCCTAGTT TTGTGCTATG
      AAGAGGTAAC AAAATCAAGA CAAAAAAACA GTCGGATCAA AACACGATAC
.....

                                HindIII
                                ~~~~~~
6551  TGTAAAAAAT ATTGCCAAGA AAAAAAGCTT GTTTTGTGGC CAGTGTCCGA
      ACATTTTTTA TAACGGTTCT TTTTTTCGAA CAAAACACCG GTCACAGGCT
.....

6601  AAAAAATTTT GGGGAATCTT CGGATTAATT TATGTTTTCA
      TTTTTTAAAA CCCCTTAGAA GCCTAATTAA ATACAAAAGT
.....
```

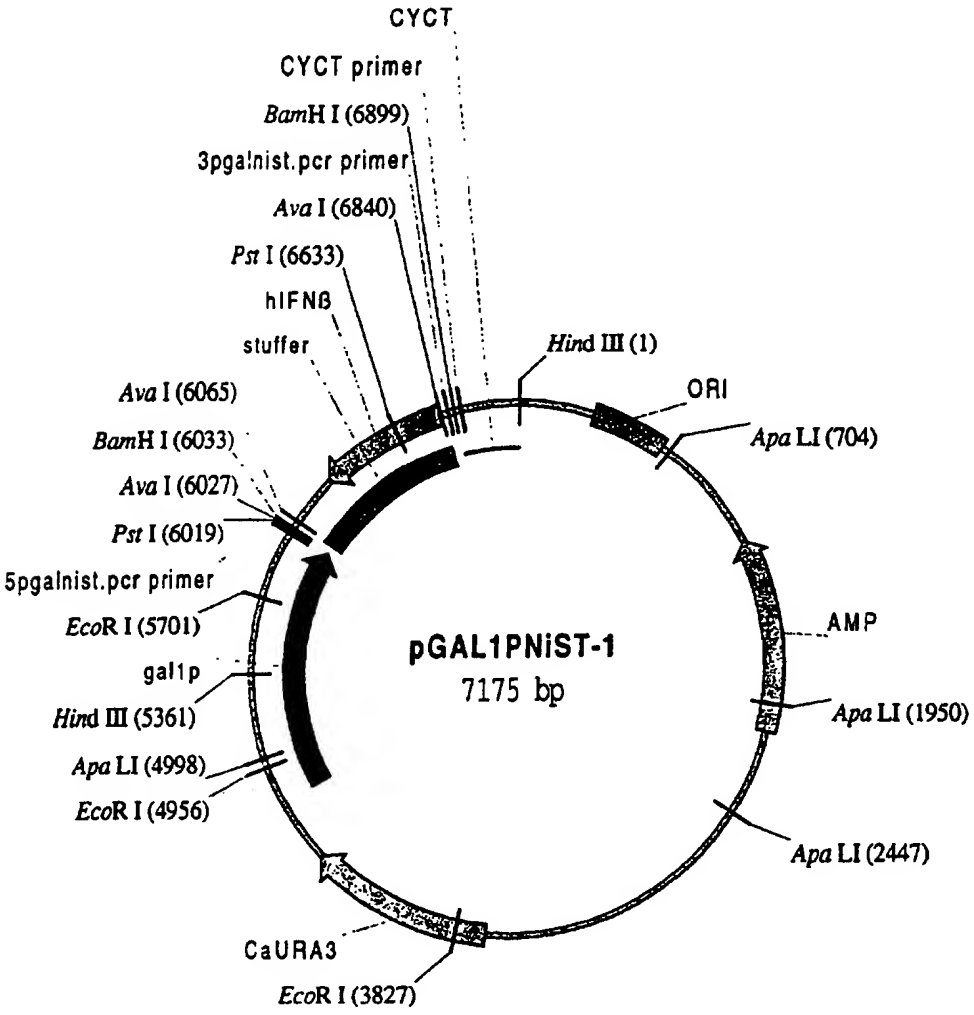

CYCT
CYCT primer
BamH I (6899)
3pgalnist.pcr primer
Ava I (6840)
Pst I (6633)
hIFNß
stuffer
Hind III (1)
ORI
Apa LI (704)
Ava I (6065)
BamH I (6033)
Ava I (6027)
Pst I (6019)
5pgalnist.pcr primer
EcoR I (5701)
gal1p
Hind III (5361)
Apa LI (4998)
EcoR I (4956)
pGAL1PNiST-1
7175 bp
AMP
Apa LI (1950)
Apa LI (2447)
CaURA3
EcoR I (3827)

Fig 64

HindIII

```
  1 AGCTTGAGTA TTCTATAGTG TCACCTAAAT AGCTTGGCGT AATCATGGTC
    TCGAACTCAT AAGATATCAC AGTGGATTTA TCGAACCGCA TTAGTACCAG
 51 ATAGCTGTTT CCTGTGTGAA ATTGTTATCC GCTCACAATT CCACACAACA
    TATCGACAAA GGACACACTT TAACAATAGG CGAGTGTTAA GGTGTGTTGT
101 TACGAGCCGG AAGCATAAAG TGTAAAGCCT GGGGTGCCTA ATGAGTGAGC
    ATGCTCGGCC TTCGTATTTC ACATTTCGGA CCCCACGGAT TACTCACTCG
151 TAACTCACAT TAATTGCGTT GCGCTCACTG CCCGCTTTCC AGTCGGGAAA
    ATTGAGTGTA ATTAACGCAA CGCGAGTGAC GGGCGAAAGG TCAGCCCTTT
201 CCTGTCGTGC CAGCTGCATT AATGAATCGG CCAACGCGCG GGGAGAGGCG
    GGACAGCACG GTCGACGTAA TTACTTAGCC GGTTGCGCGC CCCTCTCCGC
251 GTTTGCGTAT TGGGCGCTCT TCCGCTTCCT CGCTCACTGA CTCGCTGCGC
    CAAACGCATA ACCCGCGAGA AGGCGAAGGA GCGAGTGACT GAGCGACGCG
301 TCGGTCGTTC GGCTGCGGCG AGCGGTATCA GCTCACTCAA AGGCGGTAAT
    AGCCAGCAAG CCGACGCCGC TCGCCATAGT CGAGTGAGTT TCCGCCATTA
351 ACGGTTATCC ACAGAATCAG GGGATAACGC AGGAAAGAAC ATGTGAGCAA
    TGCCAATAGG TGTCTTAGTC CCCTATTGCG TCCTTTCTTG TACACTCGTT
401 AAGGCCAGCA AAAGGCCAGG AACCGTAAAA AGGCCGCGTT GCTGGCGTTT
    TTCCGGTCGT TTTCCGGTCC TTGGCATTTT TCCGGCGCAA CGACCGCAAA
451 TTCCATAGGC TCCGCCCCCC TGACGAGCAT CACAAAAATC GACGCTCAAG
    AAGGTATCCG AGGCGGGGGG ACTGCTCGTA GTGTTTTTAG CTGCGAGTTC
501 TCAGAGGTGG CGAAACCCGA CAGGACTATA AAGATACCAG GCGTTTCCCC
    AGTCTCCACC GCTTTGGGCT GTCCTGATAT TTCTATGGTC CGCAAAGGGG
551 CTGGAAGCTC CCTCGTGCGC TCTCCTGTTC CGACCCTGCC GCTTACCGGA
    GACCTTCGAG GGAGCACGCG AGAGGACAAG GCTGGGACGG CGAATGGCCT
601 TACCTGTCCG CCTTTCTCCC TTCGGGAAGC GTGGCGCTTT CTCATAGCTC
    ATGGACAGGC GGAAAGAGGG AAGCCCTTCG CACCGCGAAA GAGTATCGAG
651 ACGCTGTAGG TATCTCAGTT CGGTGTAGGT CGTTCGCTCC AAGCTGGGCT
    TGCGACATCC ATAGAGTCAA GCCACATCCA GCAAGCGAGG TTCGACCCGA
```

ApaLI

```
701 GTGTGCACGA ACCCCCCGTT CAGCCCGACC GCTGCGCCTT ATCCGGTAAC
    CACACGTGCT TGGGGGGCAA GTCGGGCTGG CGACGCGGAA TAGGCCATTG
751 TATCGTCTTG AGTCCAACCC GGTAAGACAC GACTTATCGC CACTGGCAGC
    ATAGCAGAAC TCAGGTTGGG CCATTCTGTG CTGAATAGCG GTGACCGTCG
801 AGCCACTGGT AACAGGATTA GCAGAGCGAG GTATGTAGGC GGTGCTACAG
    TCGGTGACCA TTGTCCTAAT CGTCTCGCTC CATACATCCG CCACGATGTC
851 AGTTCTTGAA GTGGTGGCCT AACTACGGCT ACACTAGAAG GACAGTATTT
    TCAAGAACTT CACCACCGGA TTGATGCCGA TGTGATCTTC CTGTCATAAA
901 GGTATCTGCG CTCTGCTGAA GCCAGTTACC TTCGGAAAAA GAGTTGGTAG
    CCATAGACGC GAGACGACTT CGGTCAATGG AAGCCTTTTT CTCAACCATC
```

Fig 65

```
 951  CTCTTGATCC GGCAAACAAA CCACCGCTGG TAGCGGTGGT TTTTTTGTTT
      GAGAACTAGG CCGTTTGTTT GGTGGCGACC ATCGCCACCA AAAAAACAAA
 .....

1001  GCAAGCAGCA GATTACGCGC AGAAAAAAAG GATCTCAAGA AGATCCTTTG
      CGTTCGTCGT CTAATGCGCG TCTTTTTTTC CTAGAGTTCT TCTAGGAAAC
 .....

1051  ATCTTTTCTA CGGGGTCTGA CGCTCAGTGG AACGAAAACT CACGTTAAGG
      TAGAAAAGAT GCCCCAGACT GCGAGTCACC TTGCTTTTGA GTGCAATTCC
 .....

1101  GATTTTGGTC ATGAGATTAT CAAAAAGGAT CTTCACCTAG ATCCTTTTAA
      CTAAAACCAG TACTCTAATA GTTTTTCCTA GAAGTGGATC TAGGAAAATT
 .....

1151  ATTAAAAATG AAGTTTTAAA TCAATCTAAA GTATATATGA GTAAACTTGG
      TAATTTTTAC TTCAAAATTT AGTTAGATTT CATATATACT CATTTGAACC
 .....

1201  TCTGACAGTT ACCAATGCTT AATCAGTGAG GCACCTATCT CAGCGATCTG
      AGACTGTCAA TGGTTACGAA TTAGTCACTC CGTGGATAGA GTCGCTAGAC
 .....

1251  TCTATTTCGT TCATCCATAG TTGCCTGACT CCCCGTCGTG TAGATAACTA
      AGATAAAGCA AGTAGGTATC AACGGACTGA GGGGCAGCAC ATCTATTGAT
 .....

1301  CGATACGGGA GGGCTTACCA TCTGGCCCCA GTGCTGCAAT GATACCGCGA
      GCTATGCCCT CCCGAATGGT AGACCGGGGT CACGACGTTA CTATGGCGCT
 .....

1351  GACCCACGCT CACCGGCTCC AGATTTATCA GCAATAAACC AGCCAGCCGG
      CTGGGTGCGA GTGGCCGAGG TCTAAATAGT CGTTATTTGG TCGGTCGGCC
 .....

1401  AAGGGCCGAG CGCAGAAGTG GTCCTGCAAC TTTATCCGCC TCCATCCAGT
      TTCCCGGCTC GCGTCTTCAC CAGGACGTTG AAATAGGCGG AGGTAGGTCA
 .....

1451  CTATTAATTG TTGCCGGGAA GCTAGAGTAA GTAGTTCGCC AGTTAATAGT
      GATAATTAAC AACGGCCCTT CGATCTCATT CATCAAGCGG TCAATTATCA
 .....

1501  TTGCGCAACG TTGTTGCCAT TGCTACAGGC ATCGTGGTGT CACGCTCGTC
      AACGCGTTGC AACAACGGTA ACGATGTCCG TAGCACCACA GTGCGAGCAG
 .....

1551  GTTTGGTATG GCTTCATTCA GCTCCGGTTC CCAACGATCA AGGCGAGTTA
      CAAACCATAC CGAAGTAAGT CGAGGCCAAG GGTTGCTAGT TCCGCTCAAT
 .....

1601  CATGATCCCC CATGTTGTGC AAAAAAGCGG TTAGCTCCTT CGGTCCTCCG
      GTACTAGGGG GTACAACACG TTTTTTCGCC AATCGAGGAA GCCAGGAGGC
 .....

1651  ATCGTTGTCA GAAGTAAGTT GGCCGCAGTG TTATCACTCA TGGTTATGGC
      TAGCAACAGT CTTCATTCAA CCGGCGTCAC AATAGTGAGT ACCAATACCG
 .....

1701  AGCACTGCAT AATTCTCTTA CTGTCATGCC ATCCGTAAGA TGCTTTTCTG
      TCGTGACGTA TTAAGAGAAT GACAGTACGG TAGGCATTCT ACGAAAAGAC
 .....

1751  TGACTGGTGA GTACTCAACC AAGTCATTCT GAGAATAGTG TATGCGGCGA
      ACTGACCACT CATGAGTTGG TTCAGTAAGA CTCTTATCAC ATACGCCGCT
 .....

1801  CCGAGTTGCT CTTGCCCGGC GTCAATACGG GATAATACCG CGCCACATAG
      GGCTCAACGA GAACGGGCCG CAGTTATGCC CTATTATGGC GCGGTGTATC
 .....

1851  CAGAACTTTA AAAGTGCTCA TCATTGGAAA ACGTTCTTCG GGGCGAAAAC
      GTCTTGAAAT TTTCACGAGT AGTAACCTTT TGCAAGAAGC CCCGCTTTTG
 .....
```

```
                                                          ApaLI
                                                          --
1901  TCTCAAGGAT CTTACCGCTG TTGAGATCCA GTTCGATGTA ACCCACTCGT
      AGAGTTCCTA GAATGGCGAC AACTCTAGGT CAAGCTACAT TGGGTGAGCA
.....

      ApaLI
      ----
1951  GCACCCAACT GATCTTCAGC ATCTTTTACT TTCACCAGCG TTTCTGGGTG
      CGTGGGTTGA CTAGAAGTCG TAGAAAATGA AAGTGGTCGC AAAGACCCAC
.....

2001  AGCAAAAACA GGAAGGCAAA ATGCCGCAAA AAAGGGAATA AGGGCGACAC
      TCGTTTTTGT CCTTCCGTTT TACGGCGTTT TTTCCCTTAT TCCCGCTGTG
.....

2051  GGAAATGTTG AATACTCATA CTCTTCCTTT TTCAATATTA TTGAAGCATT
      CCTTTACAAC TTATGAGTAT GAGAAGGAAA AAGTTATAAT AACTTCGTAA
.....

2101  TATCAGGGTT ATTGTCTCAT GAGCGGATAC ATATTTGAAT GTATTTAGAA
      ATAGTCCCAA TAACAGAGTA CTCGCCTATG TATAAACTTA CATAAATCTT
.....

2151  AAATAAACAA ATAGGGGTTC CGCGCACATT TCCCCGAAAA GTGCCACCTG
      TTTATTTGTT TATCCCCAAG GCGCGTGTAA AGGGGCTTTT CACGGTGGAC
.....

2201  ACGTCTAAGA AACCATTATT ATCATGACAT TAACCTATAA AAATAGGCGT
      TGCAGATTCT TTGGTAATAA TAGTACTGTA ATTGGATATT TTTATCCGCA
.....

2251  ATCACGAGGC CCTTTCGTCT CGCGCGTTTC GGTGATGACG GTGAAAACCT
      TAGTGCTCCG GGAAAGCAGA GCGCGCAAAG CCACTACTGC CACTTTTGGA
.....

2301  CTGACACATG CAGCTCCCGG AGACGGTCAC AGCTTGTCTG TAAGCGGATG
      GACTGTGTAC GTCGAGGGCC TCTGCCAGTG TCGAACAGAC ATTCGCCTAC
.....

2351  CCGGGAGCAG ACAAGCCCGT CAGGGCGCGT CAGCGGGTGT TGGCGGGTGT
      GGCCCTCGTC TGTTCGGGCA GTCCCGCGCA GTCGCCCACA ACCGCCCACA
.....

                                                          ApaLI
                                                          -----
2401  CGGGGCTGGC TTAACTATGC GGCATCAGAG CAGATTGTAC TGAGAGTGCA
      GCCCCGACCG AATTGATACG CCGTAGTCTC GTCTAACATG ACTCTCACGT
.....

      ApaLI
      -
2451  CCATATGCGG TGTGAAATAC CGCACAGATG CGTAAGGAGA AAATACCGCA
      GGTATACGCC ACACTTTATG GCGTGTCTAC GCATTCCTCT TTTATGGCGT
.....

2501  TCAGGCGAAA TTGTAAACGT TAATATTTTG TTAAAATTCG CGTTAAATAT
      AGTCCGCTTT AACATTTGCA ATTATAAAAC AATTTTAAGC GCAATTTATA
.....

2551  TTGTTAAATC AGCTCATTTT TTAACCAATA GGCCGAAATC GGCAAAATCC
      AACAATTTAG TCGAGTAAAA AATTGGTTAT CCGGCTTTAG CCGTTTTAGG
.....

2601  CTTATAAATC AAAAGAATAG ACCGAGATAG GGTTGAGTGT TGTTCCAGTT
      GAATATTTAG TTTTCTTATC TGGCTCTATC CCAACTCACA ACAAGGTCAA
.....

2651  TGGAACAAGA GTCCACTATT AAAGAACGTG GACTCCAACG TCAAAGGGCG
      ACCTTGTTCT CAGGTGATAA TTTCTTGCAC CTGAGGTTGC AGTTTCCCGC
.....

2701  AAAAACCGTC TATCAGGGCG ATGGCCCACT ACGTGAACCA TCACCCAAAT
      TTTTTGGCAG ATAGTCCCGC TACCGGGTGA TGCACTTGGT AGTGGGTTTA
.....

2751  CAAGTTTTTT GCGGTCGAGG TGCCGTAAAG CTCTAAATCG GAACCCTAAA
      GTTCAAAAAA CGCCAGCTCC ACGGCATTTC GAGATTTAGC CTTGGGATTT
.....
```

```
2801  GGGAGCCCCC GATTTAGAGC TTGACGGGGA AAGCCGGCGA ACGTGGCGAG
      CCCTCGGGGG CTAAATCTCG AACTGCCCCT TTCGGCCGCT TGCACCGCTC
.....
2851  AAAGGAAGGG AAGAAAGCGA AAGGAGCGGG CGCTAGGGCG CTGGCAAGTG
      TTTCCTTCCC TTCTTTCGCT TTCCTCGCCC GCGATCCCGC GACCGTTCAC
.....
2901  TAGCGGTCAC GCTGCGCGTA ACCACCACAC CCGCCGCGCT TAATGCGCCG
      ATCGCCAGTG CGACGCGCAT TGGTGGTGTG GGCGGCGCGA ATTACGCGGC
.....
2951  CTACAGGGCG CGTCCATTCG CCATTCAGGC TGCGCAACTG TTGGGAAGGG
      GATGTCCCGC GCAGGTAAGC GGTAAGTCCG ACGCGTTGAC AACCCTTCCC
.....
3001  CGATCGGTGC GGGCCTCTTC GCTATTACGC CAGCTGGCGA AAGGGGGATG
      GCTAGCCACG CCCGGAGAAG CGATAATGCG GTCGACCGCT TTCCCCCTAC
.....
3051  TGCTGCAAGG CGATTAAGTT GGGTAACGCC AGGGTTTTCC CAGTCACGAC
      ACGACGTTCC GCTAATTCAA CCCATTGCGG TCCCAAAAGG GTCAGTGCTG
.....
3101  GTTGTAAAAC GACGGCCAGT GAATTGTAAT ACGACTCACT ATAGGGCGAA
      CAACATTTTG CTGCCGGTCA CTTAACATTA TGCTGAGTGA TATCCCGCTT
.....
3151  TTGGTTTTCC AATGATGAGC ACTTTTAAAG TTCTGCTATG TGGCGCGGTA
      AACCAAAAGG TTACTACTCG TGAAAATTTC AAGACGATAC ACCGCGCCAT
.....
3201  TTATCCCGTG TTGACGCCGG GCAAGAGCAA CTCGGTCGCC GCATACACTA
      AATAGGGCAC AACTGCGGCC CGTTCTCGTT GAGCCAGCGG CGTATGTGAT
.....
3251  TTCTCAGAAT GACTTGGTTG AGTACTAATA GGAATTGATT TGGATGGTAT
      AAGAGTCTTA CTGAACCAAC TCATGATTAT CCTTAACTAA ACCTACCATA
.....
3301  AAACGGAAAC AAAAAAAAGA GCTGGTACTA CTTTCTTTAA AATTATTTTA
      TTTGCCTTTG TTTTTTTTCT CGACCATGAT GAAAGAAATT TTAATAAAAT
.....
3351  TTATTTGATT TTATTTAATA GTATATATTA TATTTTGAAC GTAGATTATT
      AATAAACTAA AATAAATTAT CATATATAAT ATAAAACTTG CATCTAATAA
.....
3401  TTGTTGAAAG TTGCTGTAGT GCCATTGATT CGTAACACTA ATTCTGTATT
      AACAACTTTC AACGACATCA CGGTAACTAA GCATTGTGAT TAAGACATAA
.....
3451  AGTCATTCCT CTTGTTTGAT AGTATCCAAA AAAACGGCTA TTTTTTTGCA
      TCAGTAAGGA GAACAAACTA TCATAGGTTT TTTTGCCGAT AAAAAAACGT
.....
3501  ATCTTATTTC CTGCATATTA TACAGATAAC ATAATGAAAG AAAAAATCTT
      TAGAATAAAG GACGTATAAT ATGTCTATTG TATTACTTTC TTTTTTAGAA
.....
3551  TTTTTTTGTT CTTCAATGAT GATTTCAACC ATTCTTTTAA ACATTGATCA
      AAAAAAACAA GAAGTTACTA CTAAAGTTGG TAAGAAAATT TGTAACTAGT
.....
3601  ATTCCTGAGC AACAACCCCA TACACACTGG TTTATATACC GCCCCTTTTA
      TAAGGACTCG TTGTTGGGGT ATGTGTGACC AAATATATGG CGGGGAAAAT
.....
3651  CAGTTGAAGA AAGAAATAGA AATAGAAATA GCAAACAAAA GATATGACAG
      GTCAACTTCT TTCTTTATCT TTATCTTTAT CGTTTGTTTT CTATACTGTC
.....
3701  TCAACACTAA GACCTATAGT GAGAGAGCAG AAACTCATGC CTCACCAGTA
      AGTTGTGATT CTGGATATCA CTCTCTCGTC TTTGAGTACG GAGTGGTCAT
.....
3751  GCACAGCGAT TATTTCGATT AATGGAACTG AAGAAAACCA ATTTATGTGC
      CGTGTCGCTA ATAAAGCTAA TTACCTTGAC TTCTTTTGGT TAAATACACG
.....
```

EcoRI
~~~~~~~

```
3801  ATCAATTGAC GTTGATACCA CTAAGGAATT CCTTGAATTA ATTGATAAAT
      TAGTTAACTG CAACTATGGT GATTCCTTAA GGAACTTAAT TAACTATTTA
 .....
3851  TAGGTCCTTA TGTATGCTTA ATCAAGACTC ATATTGATAT AATCAATGAT
      ATCCAGGAAT ACATACGAAT TAGTTCTGAG TATAACTATA TTAGTTACTA
 .....
3901  TTTTCCTATG AATCCACTAT TGAACCATTA TTAGAACTTT CACGTAAACA
      AAAAGGATAC TTAGGTGATA ACTTGGTAAT AATCTTGAAA GTGCATTTGT
 .....
3951  TCAATTTATG ATTTTTGAAG ATAGAAAATT TGCTGATATT GGTAATACCG
      AGTTAAATAC TAAAAACTTC TATCTTTTAA ACGACTATAA CCATTATGGC
 .....
4001  TAAAGAAACA ATATATTGGT GGAGTTTATA AAATTAGTAG TTGGGCAGAT
      ATTTCTTTGT TATATAACCA CCTCAAATAT TTTAATCATC AACCCGTCTA
 .....
4051  ATTACCAATG CTCATGGTGT CACTGGGAAT GGAGTGGTTG AAGGATTAAA
      TAATGGTTAC GAGTACCACA GTGACCCTTA CCTCACCAAC TTCCTAATTT
 .....
4101  ACAGGGAGCT AAAGAAACCA CCACCAACCA AGAGCCAAGA GGGTTATTGA
      TGTCCCTCGA TTTCTTTGGT GGTGGTTGGT TCTCGGTTCT CCCAATAACT
 .....
4151  TGTTAGCTGA ATTATCATCA GTGGGATCAT TAGCATATGG AGAATATTCT
      ACAATCGACT TAATAGTAGT CACCCTAGTA ATCGTATACC TCTTATAAGA
 .....
4201  CAAAAAACTG TTGAAATTGC TAAATCCGAT AAGGAATTTG TTATTGGATT
      GTTTTTTGAC AACTTTAACG ATTTAGGCTA TTCCTTAAAC AATAACCTAA
 .....
4251  TATTGCCCAA CGTGATATGG GTGGCCAAGA AGAAGGATTT GATTGGCTTA
      ATAACGGGTT GCACTATACC CACCGGTTCT TCTTCCTAAA CTAACCGAAT
 .....
4301  TTATGACACC TGGAGTTGGA TTAGATGATA AAGGTGATGG ATTAGGACAA
      AATACTGTGG ACCTCAACCT AATCTACTAT TTCCACTACC TAATCCTGTT
 .....
4351  CAATATAGAA CTGTTGATGA AGTTGTTAGC ACTGGAACTG ATATTATCAT
      GTTATATCTT GACAACTACT TCAACAATCG TGACCTTGAC TATAATAGTA
 .....
4401  TGTTGGTAGA GGATTGTTTG GTAAAGGAAG AGATCCAGAT ATTGAAGGTA
      ACAACCATCT CCTAACAAAC CATTTCCTTC TCTAGGTCTA TAACTTCCAT
 .....
4451  AAAGGTATAG AAATGCTGGT TGGAATGCTT ATTTGAAAAA GACTGGCCAA
      TTTCCATATC TTTACGACCA ACCTTACGAA TAAACTTTTT CTGACCGGTT
 .....
4501  TTATAAATGT GAAGGGGGAG ATTTTCACTT TATTAGATTT GTATATATGT
      AATATTTACA CTTCCCCCTC TAAAAGTGAA ATAATCTAAA CATATATACA
 .....
4551  AGAATAAATA AATAAATAAG TTAAATAAAT AATTAAATAA GGGTGGTAAT
      TCTTATTTAT TTATTTATTC AATTTATTTA TTAATTTATT CCCACCATTA
 .....
4601  TATTACTATT TACAATCAAA GGTGGTCCTT CTAGCTGTAA TCCGGGCAGC
      ATAATGATAA ATGTTAGTTT CCACCAGGAA GATCGACATT AGGCCCGTCG
 .....
4651  GCAACGGAAC ATTCATCAGT GTAAAAATGG AATCAATAAA GCCCTGCGCA
      CGTTGCCTTG TAAGTAGTCA CATTTTTACC TTAGTTATTT CGGGACGCGT
 .....
4701  GCGCGCAGGG TCAGCCTGAA TACGCGTTTA ATGACCAGCA CAGTCGTGAT
      CGCGCGTCCC AGTCGGACTT ATGCGCAAAT TACTGGTCGT GTCAGCACTA
 .....
```
~~~~~~~

```
4751  GGCAAGGTCA GAATAGCCCA AGTCGGCCGA GGGGCCTGTA CAGTGAGGGA
      CCGTTCCAGT CTTATCGGGT TCAGCCGGCT CCCCGGACAT GTCACTCCCT
.....
4801  AGATCTGATA TTGACGAAGA GGAACCAATG TAACGTTACA CTGAAGAAAA
      TCTAGACTAT AACTGCTTCT CCTTGGTTAC ATTGCAATGT GACTTCTTTT
.....
4851  CACACAATAA ACGGGAAGAA ACGGTGTAAA AGTGTGAAAA TAATTTTTGA
      GTGTGTTATT TGCCCTTCTT TGCCACATTT TCACACTTTT ATTAAAAACT
.....
4901  ATATCATTTC CCTTGGTTTA ATTCCAAACG AAACGTGTTT TTTTTAGAGA
      TATAGTAAAG GGAACCAAAT TAAGGTTTGC TTTGCACAAA AAAAATCTCT
.....
          EcoRI                                          ApaLI
          ~~~~~~                                         ~~~~
4951  ATGGGAATTC TTATTGGATG TCTAGATTGT TTGTTTACTC CAGACTGTGC
      TACCCTTAAG AATAACCTAC AGATCTAACA AACAAATGAG GTCTGACACG
.....
      ApaLI
      ~~
5001  ACAAAAACGT TTGGATGGAT GATCAGAAGA TATTTTTAGG CTTAGCTCTA
      TGTTTTTGCA AACCTACCTA CTAGTCTTCT ATAAAAATCC GAATCGAGAT
.....
5051  AATATAAGAA ATGATGCTTG AAAAACCAGA CAGAAATTGA GTTTCAAAAA
      TTATATTCTT TACTACGAAC TTTTTGGTCT GTCTTTAACT CAAAGTTTTT
.....
5101  TTGGTAATGT GAGGTATTAG TCAACTAACC AAATAACAAT GCAAACCGGT
      AACCATTACA CTCCATAATC AGTTGATTGG TTTATTGTTA CGTTTGGCCA
.....
5151  TGATACATTT CATTTTGAAA ATAATGAAAC TGGAATTGGA TGACCAGCAC
      ACTATGTAAA GTAAAACTTT TATTACTTTG ACCTTAACCT ACTGGTCGTG
.....
5201  ACAAACACAT AAAGTAATTA TGGGAATTAG AAGCGAACAT AGAGGAGTAC
      TGTTTGTGTA TTTCATTAAT ACCCTTAATC TTCGCTTGTA TCTCCTCATG
.....
5251  TTGGCCACGA ACAGAATACA AGTGGGAACA CTATTTTCTC CATTGTTTTA
      AACCGGTGCT TGTCTTATGT TCACCCTTGT GATAAAAGAG GTAACAAAAT
.....
5301  GTTCTGTTTT TTTGTCAGCC TAGTTTTGTG CTATGTGTAA AAAATATTGC
      CAAGACAAAA AAACAGTCGG ATCAAAACAC GATACACATT TTTTATAACG
.....
            HindIII
            ~~~~~~~
5351  CAAGAAAAAA AGCTTGTTTT GTGGCCAGTG TCCGAAAAAA ATTTTGGGGA
      GTTCTTTTTT TCGAACAAAA CACCGGTCAC AGGCTTTTTT TAAAACCCCT
.....
5401  ATCTTCGGAT TAATTTATGT TTTCATTCCA TCGGGGAAAG TGGGGGGGAA
      TAGAAGCCTA ATTAAATACA AAAGTAAGGT AGCCCCTTTC ACCCCCCCTT
.....
5451  AAAATTTTAA GCAGTTCACA AAACCTTCCA AAAAATATAT GGACAAAGAT
      TTTTAAAATT CGTCAAGTGT TTTGGAAGGT TTTTTATATA CCTGTTTCTA
.....
5501  GATTGTATTT TCCCGACACC AAAATCATAA TTAATTATGA GAAAGTTAAA
      CTAACATAAA AGGGCTGTGG TTTTAGTATT AATTAATACT CTTTCAATTT
.....
5551  TGTAACGTTA CAATTTATGT TTATTTGAAG GTGAAAAGCG ATTTATGATT
      ACATTGCAAT GTTAAATACA AATAAACTTC CACTTTTCGC TAAATACTAA
.....
5601  TTTCCGAAAT GAAAATTTTT TTTAGGTTTA TTTTTTTTGT CGGGCAAAGA
      AAAGGCTTTA CTTTTAAAAA AAATCCAAAT AAAAAAAACA GCCCGTTTCT
.....
```

```
                                                                  EcoRI
                                                                    -
5651  AAAACTGAAC AAGGATTATT AAAATTTTTG GTGTTTGTTT GTGTCTGGAG
      TTTTGACTTG TTCCTAATAA TTTTAAAAAC CACAAACAAA CACAGACCTC
 .....

      EcoRI
      -----
5701  AATTCATTCC TCTCTCATCT TCACACAATG TTTAGACATC TGACACGATT
      TTAAGTAAGG AGAGAGTAGA AGTGTGTTAC AAATCTGTAG ACTGTGCTAA
 .....

5751  CATGATAGTT CGGTTTCCGG GGTTGGTGTT TAGTTTTCGT TTTTCTTTTT
      GTACTATCAA GCCAAAGGCC CCAACCACAA ATCAAAAGCA AAAAGAAAAA
 .....

5801  TTTTGGAAAG AATGTTTTAG CTCATTGGTT TTCTTTCTTC ATTCAATAGT
      AAAACCTTTC TTACAAAATC GAGTAACCAA AAGAAAGAAG TAAGTTATCA
 .....

5851  TTTGAAAGAA TTTGCCCACT TGTTATTACA ATCATATAAA ATTAAACTTT
      AAACTTTCTT AAACGGGTGA ACAATAATGT TAGTATATTT TAATTTGAAA
 .....

5901  GATATAAAAT AGAGTTTGAA AGTTTCCCAG ATCCTTTTTG ATTTCTTTGT
      CTATATTTTA TCTCAAACTT TCAAAGGGTC TAGGAAAAAC TAAAGAAACA
 .....

5951  AAATTTTTTT TTCTCCCACA TATACACACA TACAAACCGA TTTTTATAAG
      TTTAAAAAAA AAGAGGGTGT ATATGTGTGT ATGTTTGGCT AAAAATATTC
 .....

                    PstI          AvaI    BamHI
                    ------        ------------
6001  AAAGAGTTAT ACCCTGCAGC TCGACCTCGA GGGATCCGGG CCCTCTAGAT
      TTTCTCAATA TGGGACGTCG AGCTGGAGCT CCCTAGGCCC GGGAGATCTA
 .....

                    AvaI
                    ------
6051  GCGGCCGCTA GGCCTCGAGG GACTTTTGCA CCAAAAATAA TTTATTTTCC
      CGCCGGCGAT CCGGAGCTCC CTGAAAACGT GGTTTTTATT AAATAAAAGG
 .....

6101  AAAATAAAAT TTAAATAAAT AAAAATAACT CATAATTTAA TAAAAATTTC
      TTTTATTTTA AATTTATTTA TTTTTATTGA GTATTAAATT ATTTTTAAAG
 .....

6151  AAAATCTTCT AGTGTCCTTT CATATGCAGT ACATTAGCCA TCAGTCACTT
      TTTTAGAAGA TCACAGGAAA GTATACGTCA TGTAATCGGT AGTCAGTGAA
 .....

6201  AAACAGCATC TGCTGGTTGA AGAATGCTTG AAGCAATTGT CCAGTCCCAG
      TTTGTCGTAG ACGACCAACT TCTTACGAAC TTCGTTAACA GGTCAGGGTC
 .....

6251  AGGCACAGGC TAGGAGATCT TCAGTTTCGG AGGTAACCTG TAAGTCTGTT
      TCCGTGTCCG ATCCTCTAGA AGTCAAAGCC TCCATTGGAC ATTCAGACAA
 .....

6301  AATGAAGTAA AAGTTCCTTA GGATTTCCAC TCTGACTATG GTCCAGGCAC
      TTACTTCATT TTCAAGGAAT CCTAAAGGTG AGACTGATAC CAGGTCCGTG
 .....

6351  AGTGACTGTA CTCCTTGGCC TTCAGGTAAT GCAGAATCCT CCCATAATAT
      TCACTGACAT GAGGAACCGG AAGTCCATTA CGTCTTAGGA GGGTATTATA
 .....

6401  CTTTTCAGGT GCAGACTGCT CATGAGTTTT CCCCTGGTGA AATCTTCTTT
      GAAAAGTCCA CGTCTGACGA GTACTCAAAA GGGGACCACT TTAGAAGAAA
 .....

6451  CTCCAGTTTT TCTTCCAGGA CTGTCTTCAG ATGGTTTATC TGATGATAGA
      GAGGTCAAAA AGAAGGTCCT GACAGAAGTC TACCAAATAG ACTACTATCT
 .....

6501  CATTAGCCAG GAGGTTCTCA ACAATAGTCT CATTCCAGCC AGTGCTAGAT
      GTAATCGGTC CTCCAAGAGT TGTTATCAGA GTAAGGTCGG TCACGATCTA
 .....
```

```
6551  GAATCTTGTC TGAAAATAGC AAAGATGTTC TGGAGCATCT CATAGATGGT
      CTTAGAACAG ACTTTTATCG TTTCTACAAG ACCTCGTAGA GTATCTACCA
 .....

                                   PstI
                                   -------
6601  CAATGCGGCG TCCTCCTTCT GGAACTGCTG CAGCTGCTTA ATCTCCTCAG
      GTTACGCCGC AGGAGGAAGA CCTTGACGAC GTCGACGAAT TAGAGGAGTC
 .....

6651  GGATGTCAAA GTTCATCCTG TCCTTGAGGC AGTATTCAAG CCTCCCATTC
      CCTACAGTTT CAAGTAGGAC AGGAACTCCG TCATAAGTTC GGAGGGTAAG
 .....

6701  AATTGCCACA GGAGCTTCTG ACACTGAAAA TTGCTGCTTC TTTGTAGGAA
      TTAACGGTGT CCTCGAAGAC TGTGACTTTT AACGACGAAG AAACATCCTT
 .....

6751  TCCAAGCAAG TTGTAGCTCA TGGAAAGAGC TGTAGTGGAG AAGCACAACA
      AGGTTCGTTC AACATCGAGT ACCTTTCTCG ACATCACCTC TTCGTGTTGT
 .....

                                          AvaI
                                          -------
6801  GGAGAGCAAT TTGGAGGAGA CACTTGTTGG TCATGTTCCT CGAGGCCTTT
      CCTCTCGTTA AACCTCCTCT GTGAACAACC AGTACAAGGA GCTCCGGAAA
 .....

                                                  BamHI
                                                  ---
6851  TTGGCCAGCT GGCGCCTGCT GCGCGACGGC GAGCTGCTCA CCACCCAGGA
      AACCGGTCGA CCGCGGACGA CGCGCTGCCG CTCGACGAGT GGTGGGTCCT
 .....

      BamHI
      ---
6901  TCCGTCCCCC TTTTCCTTTG TCGATATCAT GTAATTAGTT ATGTCACGCT
      AGGCAGGGGG AAAAGGAAAC AGCTATAGTA CATTAATCAA TACAGTGCGA
 .....

6951  TACATTCACG CCCTCCCCCC ACATCCGCTC TAACCGAAAA GGAAGGAGTT
      ATGTAAGTGC GGGAGGGGGG TGTAGGCGAG ATTGGCTTTT CCTTCCTCAA
 .....

7001  AGACAACCTG AAGTCTAGGT CCCTATTTAT TTTTTTATAG TTATGTTAGT
      TCTGTTGGAC TTCAGATCCA GGGATAAATA AAAAAATATC AATACAATCA
 .....

7051  ATTAAGAACG TTATTTATAT TTCAAATTTT TCTTTTTTTT CTGTACAGAC
      TAATTCTTGC AATAAATATA AAGTTTAAAA AGAAAAAAAA GACATGTCTG
 .....

7101  GCGTGTACGC ATGTAACATT ATACTGAAAA CCTTGCTTGA GAAGGTTTTG
      CGCACATGCG TACATTGTAA TATGACTTTT GGAACGAACT CTTCCAAAAC
 .....

                        HindIII
                        -
7151  GGACGCTCGA AGGCTTTAAT TTGCA
      CCTGCGAGCT TCCGAAATTA AACGT
 .....
```

328c2

```
                                      X X         fs
                                      = =         =
  1  MSITVTFPKS PSTKKRAPAF GIELEFSQQG SSDGAISKAA LAVPVFSVDN

       X X   X                                        R
       = =   =                                        =
 51  QDFVLIRDLA KYNGYPSSYQ LIVKLVKCAN IEKSQILKTD KDLNKELFEL

101  DLIBEADTKI DLFYISLPLV YSRIENKKVF YVLRESEQPK VSKAPTQEKP

151  ASVVAAESDD DNLDDDEEDE VDSDMDEDND NSGELSKGYK HMHKDNPKYI

201  NDDRVTIGQV FHQYGLDPST FLTHSLFNSI NSMSKLNYYK NFGVSGYRFL

251  PNSKLSYAER ELVLNANNYN DMHINEKTES KPKKSFRKPI GKSKKHNLQI

                              fs          T
                              =           =
301  DFNSIDLSES VIPGQGFIPD FSIHHLCKVP NYYVTSNHQS LPLSFNTKNL

                              X
                              =
351  NATSNSSYLF NDIVKIKSKS IQKLVFNSDT DNYHHTKYFY TKTYRGPGSG

401  NYKDGALMNK INKIHLSSNK KPFHKRFVSN MNRYNKSLKG LVHEKFDKNF

451  VEYLLSEQRK YTIDYSNLEI LHNSLQFNVL LNTYRGVAQE TWNNYYKFKL

                                        X           fs
                                        =           =
501  IDFEQLKALQ MEANILEERK LDAARHQQWA ESEKLRQERL RLVFEDEFNE

                              X   X     L     X   * X XX
                              =   =     =     =   = = ==
551  FEQLQSEFGQ RKKDLEEKLR RRQLEASLSD SFEADSENDD ESELAQIGQD

                                          missing  sequence
     ==========================================================
601  FESSANALKT KFEAKRYDLI NPAPPPQPIE TPQLDLNNKF SLPTVYPEIT

     missing sequence
     ==========================================================
651  RNLPLELRGV VPESXEELPP IKKAIHYVTT YPERPNIEYL TRNRDYPIAN

     missing
     ======
701  ANSGWXG
```

Fig 66

117c af

```
          L                 Q
          =                 =
   1  MSCEDEHENH MHGHIHNHNH VAPIPITAGQ SLNNKIDTSK VTALNMANSA

                ambiguities
                          =
                        Q
                        =
  51  DDLAKVFKDS TKKVPIKTK
```

Fig 67

15c1

```
                                                        fs  S
                                                        =   =
   1  QQSYVPQSQP NYSQQTQDRG MFSGGGGGHG HYQQQQGYNA YGPPPPQQGY

                                                  ambiguities
                                                       =====
                                X              W   W      WX
                                =              =   =      ==
  51  YQQQPGGGGG YYQQQQQQQP MYVQQQPRSG GNESCLMGCL AALCVCCTLD

      amb
      ==
 101  MLF
```

Fig 68

222a8

```
                    X     X         X      R        X
                    =     =         =      =        =
   1  MFDNFILKNL VDETESEIDS GECELSDDYY YYYSYEDDGK EDDSDEITAQ

  51  ILLSNSELGT KTPNFEDPFE QINIEDNKVI SVNTPKTAKP TTTVPGTSTS

        X                   XX
        =                   ==
 101  ALSTPESTIP EIPKFFVGSR RKQLSSPKNK NSTIKPDVPD WIFESGTTNE

 151  KVHGLVLVSS GVLLGTCLLF IL
```

Fig 69

222g9

```
                  E                                        E  E
                  =                                        =  =
  1  MRRREIERRK QQKXREQRQK EHEAKRDIRI QQLSEQDSRS NQTKXEEXVF

 51  KKARSTNSGA DETGLMSDKE FDDSAYSPDY LFENLWNKP NHPDTNHKTK

101  KYTENVVENL DSPPNDTSAY NSSFHDETNI QNEIQIPEND EYVPQMKATS

                              K         D     VR  fs   C
                              =         =     ===  =   =
151  SVNNTTIPAQ RRHESLSTSE NKRRXFETAD VGVXGLDSPX XAQTRNIWKI

                  P
                  =
201  QVSQNPWMVY FFMXXKRLET PEGKLLCRDQ
```

Fig 70

328c1

```
                F
                =
           ambiguities
           ==
  1  MPRIKQVDVF TNYKYLGNPV AVIYDSDNLT TQEMQKIARW TNLSETTFIL

 51  TPKSSIADYS IEIFTSGGNE LPFAGHPTLG TAFALLEDGK IKPNDNGQII

                     ambiguities
                       ========
101  QECGAGLVKI SVEKTPNNNN NNENNNKSNEL PFLLSFELPY FKFHEIDDKV

                     fs                                G    N
                     =                                 =    =
151  IEELQHSWNG TNIIGKPVLI DAGPKWAVFQ LGSGKEVLDL NVDLAQIERL

                                 ambiguity

                  ambiguities  Q         G
                  ==========   =         =
201  SLENGWTGIG VFGKHNENGD SVELRNIAPA VGV
```

Fig 71

33qK part1

```
                                    XX    XXXX                  fs
                                    ==    ======                =
  1  NVTDSTPFAM GTLGSTFYAV TSVGPSFQIY DLATLHLLFV SQTQTPSRIT

     C      XF      X
     =      ==      =
 51  SLAAHHHYVY ASYGDRIGIF RRGRLEHELV CEGNSTVNQL LVFGEYLIAT

                         X     X      XX             X      XXXX
                         =     =      ==             =      ====
101  TLEGDIFVFR KTEGKKFPTE LYTTIRIINS LVEGEIVGLI HPPTYLNKVI

     XXX X
     === =
151  VATTQSVFVI NVRTGKLLYK SRELQFEGEK ISSIEAAPVL DVIAVGTSNG

                    fs
                     =
               X     P
               =     =
201  NVFLFNIKKG ESVGAKKLLL LELNLPSKVA SISFRTDGAP HLVAGLNNGD

                    A         I        R
                    =         =        =
251  LYFYDLDKKS RVHVLRNAEK ETHGGVANAK FLNGQPIVLS NGGDNHLKEF

       H   S                 Y          G   R             C
       =   =                 =          =   =             =
301  VFDPNLTTSN SSIVPPPRHL RSRGGHSAPF VAIEFPQEDK THFLLSASRD

           fs
            =
           Y          LC                  ambiguities
           =          ==          ==========================
351  KTFWTFSLRK DAQAQEMSQR LQKSKDGKRQ AGQVVSMREK FPEIISISSS

     amb
     ==
401  YAR
```

Fig 72

33gK_part2

```
   1  IITAHKDETF ARTWDSRNKR VGRHLLNTID GGIVKSVCVS QCGNFGLVGS
  51  SLGGIGSYNL QSGLLRKKYV LHKQAVTGLA IDGMIRKMVS CGLDGIVGFY
 101  DFGKSVYLGK LQLEAPITSM IYHKLSDLVA CALDDLSIVV IDVTPQKVIR
 151  ILYGHTNRIS GMDFSPDGRW IVSVALDSTL RTWDLPTGGC IDGVILPIVA
 201  TAVKFSPIGD ILATTHVSGN GVSLWTNFAQ FKPVSTRHVE EDEFSTILLP
 251  NASGDGGSTN LDGFLDEDSN EDGTIDEQYT SAAQIDASLI TLGSEPRSKF
 301  NTLLKLDTIK QQSKPKEAPK KPENAPFFLQ LTG
```

Fig 73

33gK_part3

```
   1  KLRKLDTNGN HAFESEFTKL LREAGESGQF ERFLTYLLNL SPAVLDLEIR
  51  SLNSFVPLTE MTNFIQALNA GLKSNANYEI METLYAMFFN IHGDVIHQFE
 101  NETSLHEALE EYRQLNDEKI NKMDSLVKYC ASIVSFIS
```

Fig 74

480c

```
          X    X                    X     X         X
          =    =                    =     =         =
   1  MLQNNRADLD XVQNDWYDVY RKLKLDILNE SSSSISEQIH IRDRISRVYQ

       fsG X       Xfs      XG  FFXXXE  V    S       DPTX
        ==  =       = =      ==  =======  =    =       ====
  51  PRILDLVRAI GTDKEEALKK KQLVSQLQES IDNLLVQEVP RSKRVLGGAV

      *XFGXXXXIXQ*X   X  INV                missing sequence
      ==============  =  ================================
 101  KETPETLPLN NKELLQHQVQ IHQNQDKELD QLRVLIARQK QIGELINAEV

      missing sequence
      ================================
 151  EEQNEMLDRF NEEVDYTSSK IKQARRRAKK IL
```

Fig 75

55q1

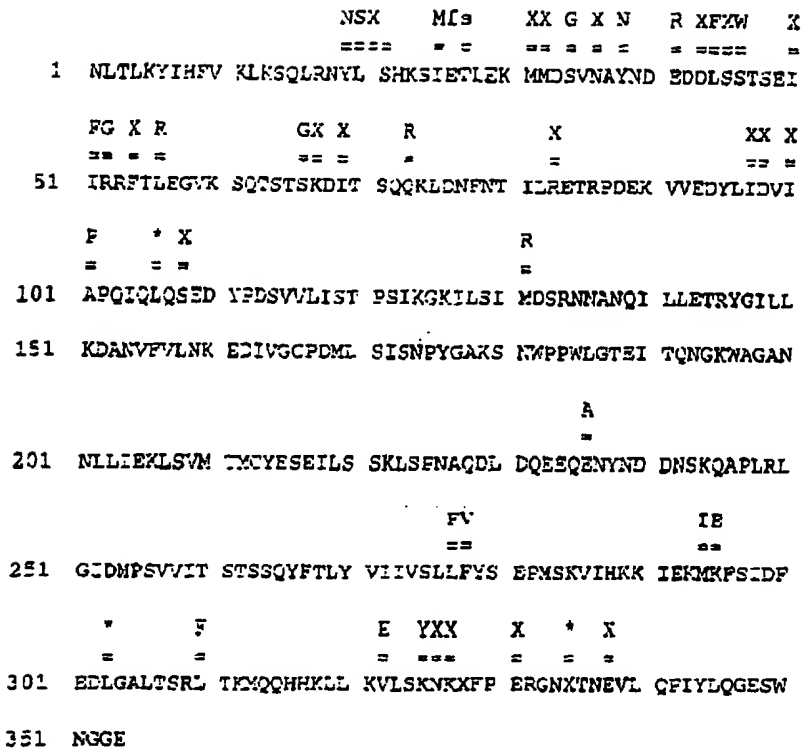

NSX MIs XX G X N R XFXW X
1 NLTLKYIHFV KLKSQLRNYL SHKSIETLEK MMDSVNAYND EDDLSSTSEI
FG X R GX X R X XX X
51 IRRFTLEGVK SQTSTSKDIT SQQKLDNFNT ILRETRPDEK VVEDYLIDVI
P * X R
101 APQIQLQSED YPDSVVLIST PSIKGKILSI MDSRNNANQI LLETRYGILL
151 KDANVFVLNK EDIVGCPDML SISNPYGAKS NWPPWLGTEI TQNGKWAGAN
A
201 NLLIEKLSVM TMCYESEILS SKLSFNAQDL DQESQENYND DNSKQAPLRL
FV IB
251 GIDMPSVVIT STSSQYFTLY VIIVSLLFYS EPMSKVIHKK IEKMKPSIDF
* F E YXX X * X
301 EDLGALTSRL TKMQQHHKLL KVLSKNKXFP ERGNXTNEVL QPIYLQGESW
351 NGGE

Fig 76

60gK

```
                              FIG
                              ===
  1  ITDFSDFKTT KLPALAELDI LKRCYICKDL LNAPVRTQCD HTYCSQCIRE

                                                    *
                                                    =
 51  FILRDNRCPL CKTEVFESGL KRDFLLEEIV ISYASLRPHL LRLLEIEKVE

101  SKQEVDREKS ANESALNGNR NVNNDVDETV RVKDQLNADK LGEEKGQAQH

     G  fs                                              X
     =  =                                               =
151  WEQVNEQTTE VILLLSDDEE NGSDSLVKCP ICFERMELDV LQGKHIDDCL

                                                      fs
                                                      =
                          Q          Q      X ambiguities
                          =          =      =    =====
201  SGKSTKRTPT DILSPKAKRP KQITSFFKPT IDTKTPSPPT SKASTTPTAT

                 S  Q       N          I     K         M
                 =  =       =          =     =         =
251  PTTTLLKALV ASPSPVAQST VHKGKPLPKL DFSSLSTQKI KAKLSDLKLP

                   *          T
                   =          =
301  TTGSRNEMEA RYLHYYTTVN ANLDSNHPV
```

Fig 77

8c gp

```
                                                      fs
                                                      =
       G                        G     fs        D     A
       =                        =     =         =     =
  1  MQFSSAVVLS AVAGSALAAY SNSTVTDIQT TVVTITSCEE NKCHETEVTT

 51  GVTTVTEVDT TYTTYCPLST TEAPAPSTAT DVSTTVVTIT SCEEDKCHET

101  AVTTGVTTVT EGTTIYTTYC PLPSTEAPGP APSTAEESKP AESSPVPTTA

151  AESSPAKTTA AESSPAQETT PKTVAAESSS AETTAPAVST AEAGAAAMAV

201  PVAAGLLALA ALF
```

Fig 78

29q3 part1

```
          X  X
          =  =
   1  REESIISEEM KEGIPFFTIV ARIPVIEAFG FSEDIRKKTS GAASPQLVFD

  51  GYDMLDIDPF NVPHIEEELE ELGEFAEREN VARRYMMNIR RRKGLFVDEK

          X    X
          =    =
 101  VVKNAEKQRT IKRD
```

Fig 79

29q3 part2

```
   1  VMRLQQGSQE PEVHEHLINL IDSPGHIDFS SEVSTSSRLC DGAVVLVDVV

  51  EGVCSQTVTV LRQCWIDKLK PLLVINKIDR LITEWKLSPL EAYQHISRII

 101  EQVNSVIGSF FAGDRLEDDL NWREAGSVGE FIEKSDEDLY FTPEKNNVIF

 151  ASAIDGWAFS VNQFAKIYLK KLGFSQQALS KTLWGDFYLD MKNKKIIPGK

 201  KLKNNSNSLK PLFVSLILDQ VWAVYENCVI ERNQDKLEKI IEKLGAKITP

                                                          XXX
                                                          ===
 251  KDLRSKDYKN LENLIMSQWI PLSHAILGSV IEYLPSPIVA QRERIDKILD

      XXX XXXX XXXX      XX   X          ambiguities
      === ==== =====     ==   = ==========**==========
 301  ETIYSAVDSE LEKSKLVDPS FVKAMQECDS SHPETHTIAY VSKLLSIPNS

           fs                                 ambiguities
           =                                    ===
 351  DLPKASNAAT GGLTADEIQE RGRIARELAK KASEAAALAQ EGSKNEDEFA

        X
        =
 401  IKPKYDPFEW EFEEDDFENS EDESDANAVE ESTETIVGFT RIYSGSLSRG

                    X            S          Y   G        NFR
                    =            =          =   =        ===
 451  QKLTVIGPKY DPSLPRDHQT NFEQITNEVE IKDLFLIMGR ELVRMEKVPA

                    F          *R         V
                    =          **         =
 501  GNIVGVVGLD NAVLKNATIC SPLPEDKPYI NLASTSTLIH NKPIMKIAVE

 551  PTNPIKLAKL ERGL
```

Fig 80

17c cp

```
  1  PPKVARSXSS TIGKIFRYTF YTAVISVIGS AGLIGYRIYE ESQPVDQVKQ

                                          X
                                          =
 51  TPLFPNGEKK NILVILGSGW GAISLLKNLD TTLYNVVIVS PRNYFLFTPL

                                fs        X  fs    fs
                                =         =  =     =
101  LPSVPTGTVE LRSIIEPVRS VTRRCPGQVI YLEASATNIN PKTNELTLKQ

       R     N                                      X X
       =     =                                      = =
151  STTVVSGHSG KDTSSSKSTV AEYTGVESIT TTLNYDYLVV GVGAQTILIF

       X X X       XX   XX X
       = = =       ==   == =
201  GNPGPRMRKF NPFFERTTSG SHLQIR
```

Fig 81

357cL

```
  1  MAKFIKAGKV AIVVRGFVAG KKVVIVKHHD EGTKSHPFPH AIVAGIERAP

 51  LKVTKKMDAK KVTKRTKVKP FVKLVNYNHL MPTRYSLDVE SFKSAVTSEA

             fs
             =
101  LEEPSQREEA KKVVKKAFEE KHQAGKIWWF FQKLHF
```

Fig 82

110c_gf

```
                                   *
   1 MSIPSTQVGF FVKKASGLNL KKDLPDNKPG AGQLLLKVDA VGLCHSDLHV

                                                        fs
                                                        =
                                  X  fs       D     X  X X
                                  =  =        =     =  = =
  51 LYEGLDCGDN VVMGHEIAGT VAELGEEVSE FAVGERVACV GPNGCGLCKH

                                                missing  sequence
                                                ==========
          ambiguities   X     XXRXIPGXXEXX
          ============  =     ============
 101 CLTGNDNVCT KSFLDWFSLG YNGGYEQFLL VKRPRNLVKI PDNVTSEEAA

                                                missing  sequence
     ==========================================================
 151 AITDAVLTPY HAIKSAGVGP ASNILIIGAG GLGGNAIQVA KAFGAKVTVL

                                             missing sequence
     ====================================================
 201 DKKDKARDQA KAFGADQVVS ELPDSVLPGS FSACFDFVSV QATYDLCQKY

     missing sequence
     ================================================
 251 CEPKGTIVFV SLGAISLNIN LALLDLREIT VKGSFWGTLM DLREAFELAA

     missing sequence
     ================================
 301 QGKVKPNVAH APLSELPKYM EKLRAGGYEG RVVFNP
```

Fig 83

409c10

```
          M  Y   N            E             D
          =  =   =            =             =
   1  MEKIDINTNS NKIQQAYDKV VRGDENATFV VYSVDKNATM DVTSTGDGSL

      Y                P                            P   LSYD
      =                =                            =   ====
  51  EDFVEHFTDG QVQFGLARVT VPGSDVSKNI LLGWCPDSAP AKLRCAFANN

      S     I          M   N  Y  H     X
      =     =          =   =  =  =     =
 101  FADVSRVLSG YHVQITARDQ DDLDVNEFLN RVGAAAGARY STQTSGLKKP

 151  SPAAPKPTSK FVVAKPSSAS KPSFVPKSTG KPVAPAKPKP KNITKDAGNG

 201  DAEDVEERDP DKKPLDNVPS AYKPTKVNID ELPKQKSDTT SSTPKTFKSE

               P             HD  R           T
               =             ==  =           =
 251  PQEEKNDDDG QSKPLSERMK AYDQPSSSDG RLTSLPKPKI GHSVADKYKA

                        fs
                        =
 301  SASGNGAAPA FGAKPAFGIQ SVDSRKDKLV GGLSRDFGAE NGKTPAQIWA

 351  EKRGKYKTVA SDEKECNSSE KVDEPSEHHA ADLAKKFEEK ANIAGDTPSL

                               K          S   fs        X
                               =          =   =         =
 401  PTRNLPPAPP ARETAIPSNE KQKEEKEEEE QAPAPSLPTR NLPPPSQRQP

      fs   X            XS                 missing  sequence
       =   =            ==          ----------------------
 451  EPEPEPEEEE EEEEEEAPAP SLPARNLPPA PKAEAEESKK QSTTATAEYD

          missing  sequence
      ------------------------------------------------------
 501  YEKDEDNEIG FSEGDLIIDI EFVDDDWWQG KHAKTGEVGL FPATYVSL
```

Fig 84

226c #2

```
   1  VLGSQWGDEG KGKLVDLLCD DIDVCARCQG GNNAGHTIVV GKVKYDFHML

  51  PSGLVNPKCQ NLVGSGVVIH VPSFFAELEN LEAKGLDCRD RLFVSSRAHL

          A     D   IX
          =     =   ==
 101  VFDFHQRTDK LKEAELSTNK KSIGTTGKGI GPTYSTKASR SGIRVHHLVN

 151  PDPEAWEEFK TRYLRLVESR QERYGEFEYD
```

Fig 85

CDC48

```
                                        P
                                        =
  1 EDKKQHFDAS GASAVDDKTA TALLRRKKKE NALVVDDATN DDNSVITMSS

 51 NTMELLQLFR GDTVLVEGKK RKDTVLIVLA DDDMPDGVAR VNRCVRNNLR

101 VRLGDIVTVH PCPDIKYANR ISVLPIADTV EGINGSLFDL YLKPYFMEAY

151 RPVRKGDLFT VRGGMRQVEF KVVEVDPEEI AIVAQDTIIH CEGEPINRED

                    R   fs
                    =   =
201 EENSLNEVGY DDIGGCKKQM AQIRELVELP LRHPQLFKSI GIKPPKGILM

251 YGPPGTGKTL MARAVANETG AFFFLINGPE IMSKMAGESE SNLRKAFEEA

                                    fs        P  fs     N KK  P
                                    =         =  =      = == =
301 EKNSPSIIFI DEIDSIAPKR DKTNGEVERR VVSQLLTLMD GMKARSNVVV

      G     DQ      L Pfs
      =     ==      = ====
351 IAATNRPNSI DPALRRFGRF DREVDIGVPD AEGRLEILRI HTKNMKLADD

401 VDLEAIASET HGFVGADIAS LCSEAAMQQI RERMDLIDLE EETIDTEVLN

451 SLGVTQDNFR FALGNSNPSA LRETVVENVN WTWDDIGGLE NIKNELKETV

501 EYPVLHPDQY QKFGLAPTKG VLFFGPPGTG KTLLAKAVAT EVSANFISVK

551 GPELLSMWVG ESESNIRDIF DKARAAAPTV VFLDELDSIA KARGGSHGDA

601 GGASDRVVNQ LLTEMDGMNA KKNVFVIGAT NRPDQIDPAL LRPGRLDQLI

651 YVPLPDEPAR LSILQAQLRN TPLEPGLDLK EIAKITHGFS GADLSYIVQR

                          AfsQ I
                          = == =
701 SAKFAIKDSI EAQVKINKIK EEKEKVKTED VDMKVDEVEE EDPVPYIIRA

751 HFEEAMKTAK RSVSDAELRR YESYAQQLQA SRGQFSSFRF NENAGATDNG

                X
                =
801 SAAGANSQAA FQX
```

Fig 86

99a3 part1

```
                                        fs
                                        =
  1  TLKQRLEEIL FAKAEEVKQF KKEHGKIVIG FVLLEQAYGG MRGIKGLVWE

 51  GSVLDPIEGI RFRGRTIPDI QKELPKAPGG EEPLPEALFW LLLTGEVPTD

101  AQTKALSEEF AAPSALPKHV EELIDRSPSH LHPMAQFSIA VTALESESQF

151  AQAYAKGANK SEIWKYTYED SIDLLAKLPT IAAKIYRNVF HDGKLPAAID

                X     T   X                              T
                =     =   =                              =
201  SKLDYGANLA SLLGFGDNKE FVELMRLYLT IHSDHEGGNV SAHTTHLVGS

                     *      RI   *
                     =      ==   =
251  ALSSPFLSLA AGLNGLAGPL HGRANQEVL
```

Fig 87

99a3 part2

```
  1  QREFALKHMP DYELFKLVSN IYEVAPGVLT KHGKTKNPWP NVDSHSGVLL

                       S
                       =
 51  QYYGLTEQSF YTVLFGVFRA FGVLPQLILD RGIGMPIERP KSFSTEKYIE

101  LVKNINK
```

Fig 88

409c5 part1

```
     ambiguityX                X          *  GL
     =         =               =          =  ==
  1  SDYHVIWLAK RNNGIMEAEY RLYLLVITLI ISPVSXIMFG VGAAREWPWQ

 51  VIYVGLGFIG FGWGSIGDTS MSYLMDAYPD IVIQGMVGVS IINNTLACIF

101  TFACSYWLDG SGTQNTYIAL SIIDFATIAL VFPFLYYGKT FRRKTKRLYV

151  SMVELTQGM
```

Fig 89

409c5 part2

```
  1  DQNNEDFIPG TLNIVSLEVD SEDENVGHYD ASSRPKVKTK GNIILFPQPS
 51  NSCNDPLNWS KNRKLSNFFI VIPITAFTAA TSNDAGSIQD SLNEKYGISY
101  DAMNTGAGVL FLGIGWGTFF LTPASSLYGR KITYFICIFL GLLGAVWFAL
151  VKSISDSINS QLFVGISESC AEAQVQLSLS ELYFAHNLGS VLTSYIVATS
201  VGTYLGPLIA AFIVQNIGFR WVGWIAAIIS GALLFVIVFC LDETYFDRAK
251  FTKP
```

Fig 90

ESP1

```
  1  DFQLQDILHH VESKWFGGFI SGIFTNDNDV ENESKNVFHK FKQDLMKILK
 51  DCLTVSDDKS NIEPFLQFNE FIYYCFYSME EYNYELVDDL IKFITINMNS
                    X                       *
                    =                       =
101  HGRIVNFGTN VKINKLNELI KNLIDKVNKN KQRCDKQQQK QQQQQQQQQQ

     ambiguities
       ===
     QQQQ  fs        X          ambiguities
     ====  =         =     ========================
151  NSNNSQHIVL IPNANCSNFP WESMEFLRSK SISRMPSIHM LLDLVKSNTN

201  NKNKLMFVDK SNLYYLINPS GDLIRSENRF KKLFESNHLW RGEIGKLSSN

                                                    F
                                                    =
                                       X      ambiguities
                                       =      ========
251  EHEDYQDSIL CEILKSHLFV YIGHGGCDQY IKVSKLFKKC GNNQDLLNKL

301  PPSLLLGCSS VKLDNCNYNY NSSMLQPLGN IYNWLNCKSS MILGNLWDVT

                         D
                         =
351  DKDIDIFTLS LLQKWGLIAD YNGSGHDYGM KKLDLTNCVV QSRSKCTLKY

401  LNGSAPVVYG LPK
```

Fig 91

190g3

```
                                        A
                                        =
  1  STKGIKVHRT FESMNLKPDL LKGIYAYGFE TPSAIQSRAI MQIISGRDTI

                                                          M
                                                          =
 51  AQAQSGTGKT ATFSIGMLEV IDTKSKECQA LILSPTRELA IQIQNVVKHL

101  GDYMNIHTHA CIGGHNVGED VKKLQQGCQI VSGTPGRVID VIKRRNLQTR

151  NIKVLILDEA DELFTKGFKE QIYEIYKHLP PSVQVVVVSA TLPREVLEMT

     fs                      T
      =                      =
201  SKFTTDPVKI LVKRDEISLL GIKQYYVQCE REDWKFDTLC DLYDNLTITQ

251  AVIFCNTKLK VNWLADQMKK QNFTVVAMHG DMKQDERDSI MNDFRRGNSR

301  VLISTDVWAR GIDVQQVSLV INYDLPTDKE NYIHRIGRSG RFGRKGTAIN

                        X
                        =
351  LITKDDVVTL YELEKYYSTK IKEMPMNIND IM
```

Fig 92

40c af

```
  1  NVDIDEITLT RFILQEGQTV APTATGELSL LLNALQFAFK FIAHNIRRAE

 51  LVNLIGVSGS ANSTGDVQNN LIVIGDEIFI NAHRSSNNVK VLVSEEQEDL

101  IVFPGGGTYA VGTFPIDGSS NIDAGVSVGT IFGVYKLQEG STGGISDVLR

                                        missing sequence
                                        ================
151  PGKEMVAAGY TMYGASAHLA LTTGHGVNLF TLDTQLGEFI LTHPNLKLPD

                                        missing sequence
     ==========================================================
201  TKNIYSLNEG YSNKFPEYVQ DYLKDIKKEG YSLRYIGIMV ADVHFTLLYG

     missing sequence
     ================================================
251  GIFAYPTLKL RVLYEIFPMA LLMEQAGGSA VTIKGERILD ILPKGIHDKS

     missing sequence
     ================
301  SIVLGSKGEV EKVLKHVPK
```

Fig 93

360c6

```
   1  DNVSSTSTAE AVNNEIKVKD EFPQEESQAH T SLEDKPVSAY IGIIIMCFLI

  51  AFGGFVFGFD TGTISGFINM SDFLERFGGT KADGTLYFSN VRTGLMIGLF

            X X                 X
            = =                 =
 101  NAGCAIGALF LSKVGDMYGR RVGINTAMIV YIVGIIVQIA SQHAWYQVMI

                                                 ambiguities
                                                    ==
                                                  X
                                                  =
 151  GRIITGLAVG NLSVLCPLFI SEVSPKHLRG TLVCCFQLMI TLGIFLGYCT

      fs
      =
 201  TYGIKSYSDS RQWRIPLGLC FAWALCLVAG MVRMPESPRY LVGKDRIEDA

                                                   PR
                                                   ==
 251  KMSLAKTNKV SPEDPALYRE LQLIQAGVER ERLAGKASWG TLFTGKTKIF

         IV          missing  sequence
         ==          ==================
 301  ERVMLGVMLQ ALQQFNWGKN LFPSYLTSXP N
```

Fig 94

98c cp

```
                                              missing  sequence
      ========================================================
  1   NAPVSGTITE FLVDVDATVE VGQEIIKMEE GDAPAGGASA SEAPAKKEEA

                                              missing  sequence
      ========================================================
 51   PEKAKEESAP AAAPKKEETK KEEPKKESKP APKKEESKKS TQSTTSAPTF

          missing  sequence
      ====================================================
101   TNFSRNEERV KMIRVRLRIA ERLKESQNTA ASLTTFNEVD MSNLMDFRKK

      missing sequence
      ================================================
151   YKDEFIEKTG IKLGFMGAFS KASALALKEI FAVNAAIENN DTLVFKDYAD

      missing sequence              X   XX  XX   NX     *
      ============================= =   ==  ==   ===    =
201   ISIAVATPKG LVTPVVRNAE SLSILGIEKE ISNLGKKARD GKLTLEDMTG

       S   X XX  X    C     X X* X F    XF X  IX
       =   = ==  =    =     = == = =    == =  ==
251   GTFTISNGGV FGSLYGTPII NMPQTAVLGL HGVKERPVTV NGQIVSRPMM

301   YLALTYDHRV VDGRESVSFL RTIKELIEDP RKMLL
```

Fig 95

249c of

```
   1  ERMISGMIYN CLQKELETTR MSCRDYMLDY GSFRTRDYKT TQEFLDAKYK

  51  HLESPIGHVG KQAFMEYPIY FDYGFNTYLG DNFYSNYNLT ILDVSIVRIG

                                                          fs
                                                          =
 101  NNVKCGPNYS ILIPTHPVDP TLRYDQLENA LPVTVGNGVN LCGSCTILGG

        X              missing  sequence
        =    ==========================================
 151  VTVGDGSIVA AGAVVNKDVP PNTVVAGVPA RVVKQLEPRD PNF
```

Fig 96

55a1

```
        XX   XXPY     X   P   X      X
        ==   ====    =   =   =      =
   1  TSDTKTKQRR NPLLKDISSQ GGNLRTVPRS SSSGSSQKKK SSKKQRHNDE

                                                T  HL   V  L
                                                =  ==   =  =
  51  DDEENGGGEG PLDASSERKI LQLAKEQQDE LEQEDEIQNK PSFAQSFKNQ

                    D   L          L        FI             S
                    =   =          =        ==             =
 101  QIDSEEEEEE DEYSDTEEEE EVEEIVVDEE DAEVDPKDAE LFNKYFQSNG

       D                    FX   R X   P  X    F
       =                    ==   = =   =  =    =
 151  EANNNDDDNS FQPTINLADK ILAKIQEKES QQQQQQQSSP DNSNEDAVLL

              E          I
              =          =
 201  PPKVILAYKK IGQILSTYTH GKLPKLFKIL PSLKNWQDVL YVTNPNSWTP

 251  HATYEATKLF VSNLSSNEAT VFIETILLPR FRDSIENSDD HSLNYHIYRA

 301  LKKSLYKPGA FPKGFLLPLV DGYCSVREAT IAASVLTKVS VPVLHSCHYC

 351  GVLMNKKRES PVFVLRRI
```

Fig 97

```
  1 MYVYKRDGRK EPVRFDKITA RVQRLCYGLN PNHVEPVAIT QKVISGVYQG
 51 VTTIELDNLA AETAATMTTI HPDYAVLAAR IAVSNLHKQT TKQYSKVSKD
101 LYEYINPKTG LHSPMISKET YDIIMEHEDE LNSAIVYDRD FNYNYFGFKT
151 LERSYLLRIN GKVAERPQHL IMRVAVGIHG NDIPRVIETY NLMSQRFFTH
201 GSPCLFNAGT PRPQMSSCFL LAMKDDSIEG IYDTLKSCAL ISKSAGGIGL
251 HIHNIRSTGA YIAGTNGTSN GIIPMVRVFN NTARYVDQGG NKRPGAFALY
301 LEPWHSDIFD FIDIRKNHGK EEIRARDLFP ALWIPDLFMK RVEQNGDWTL
351 FSPNEAPGLA DVYGDEFEEL YTKYEKENRG RQTIKAQKLW YAILGAQTET
401 GTPFMLYKDS CNNKSNQKNL GIIKSSNLCC EIVEYSAPDE VAVCNLASIA
451 LPSFVENDEK STNYNFDKLH QVTKVVTRNL NRVIDRNHYP VPEAERSNMR
501 HRPIALGVQG LADAFMELRL PFDSQEAREL NIQIFETIYH AAVEASIELA
551 KEEGAYETYP GSPASQGLLQ FDLWNRKPTE LWDWDTLKQD LAKHGMRNSL
601 LVAPMPTAST SQILGNNECF EPYTSNIYSR RVLAGEFQIV NPYLLKDLVD
651 LGVWNDAMKS SIIANNGSIQ ALPNIPDEIK ALYKTVWEIS QKHIIDMAAD
701 RAAFIDQSQS LNIHIKDPTM GKLTSMHFYG WKKGLKTGMY YLRTQAASAA
751 IQFTIDQKIA ETAGHIVAML DKLNIKKYVN KGRVESENTS DAPYKSFSTE
801 PTSLESSVAD LKIKDEGEKP AEDKTIEELE NDIYSAKVIA CAIDNPESCT
851 MCSG
```

Fig 98

485cl

```
          fs    X    fs          V
           =    =    =          =
  1 APKWYQSEDV PAPKQTRKTA RPQKLRASLA PGTVLILLAG RFRGKRVVYL
 51 KNLEDNTLLV SGFFKVNGVP LRPVNARYVI ATSTKVNVSG VDVSKFNVEY
101 FAREKSSKSK KSEAEFFNES QPKKEIKAER VADQKSVDAA LLSEIKKTPL
151 LKQYLAASFS LKNGIRFHLL KF
```

Fig 99

328c3

```
   1  MENDKGQLVE LYVPRKCSAT NRIIKAKDHA SVQISIAKVD EDGRAIAGEN

  51  ITYALSGYVR GRGEADDSLN RLAQGDGLLK NVWSYSR
```

Fig 100

67g1_part1

```
                                                          X
                                                          =
   1  MSFKGFKKGV LRAPQTNRQK FNMGEITQDA VYLDAERRFK EIEMETKKLS

  51  EESKKYFNAV NGMLDEQIDF ARAVAEIYKP ISGRLSDPSA TVPEDNPQGI

                                        P
                                        =
 101  EASELYQAVV KDLKDTLKPD LELIEKRIVE LAQELLKIIQ AIRKMS
```

Fig 101

67g1_part2

```
                                  A
                                  =
   1  HKRNFSKYEL KKERTVKDEE KMFSAQTEVE IAQQEYDVYN DLLKNELPVL

      Y           Y                                 I       D
      =           =                                 =       =
  51  FQMQSDFIKP LFVSFYYMQL NIFYTLYTRM ELKIPYFDL STDIVEAYTA

                S                                 fs      fs M
                =                                 =       = =
 101  KKGNIESQTD AIGITHFKVG HAKSKLEATK RRHAAMNSPP PTGASSIAST

 151  GTGGELPAYS PGGYNQPYGD SKYQPPSSPA TYQSPVVAAT AQSPATYQSP

 201  VATGQPPSYL PQTPASAPPP QVGSGLPTCT ALYDYTAQAQ GDLTFPAGAV

 251  IEIIQRTEDA NGWWTGKYNG QTGVFPGNYV QL
```

Fig 102

```
  1 MTTSXETFLF TSESVGEGHP DKICDQVSDA ILDACLAVDP LSKVACETAA
 51 KTGMIMVFGE ITTKAQLDYQ KIIRDTIKHI GYDDSEKGFD YKTCNVLVAI
101 EQQSPDIAQG LHYEKALEEL GAGDQGIMFG YATDETDEKL PLTILLAHKL
151 NAALASARRS GSLPWLRPDT KTQVTIEYEK DGGAVIPKRV DTIVISTQHA
201 EEITTENLRK EIIEHIIKQV IPEHLLDDKT IYHIQPSGRF VIGGPQGDAG
251 LTGRKIIVDT YGGWGAHGGG AFSGKDFSKV DRSAAYAARW VAKSLVTAGL
301 AKRALVQFSY AIGVAEPTSI YIDTYGTSKL STEALVEIIK NNFDLRPGVI
351 VKELDLARPI YFKTASYGHF TNQENSWEQP KKLKF
```

Fig 103

232e CR

```
  1 FKNSADAKYT MEILRGIRLY GKALKLKRID AKSQSSTNNP NNQTIGIFVQ
 51 SDLINPNYID VGAKLFINNL NPLVDESFLM DTFSKFGTLI RNPIIRRDSE
                                    T missing sequence
                                    =          ==========
101 GHSLGYGFLT YDDFESSDLC IQKGNNTILM NNKIAISYAF KDLSVDGKKS

    missing sequenc
    ===============
151 RHGDQVERKL GKKCNP
```

Fig 104

83c3

```
                ambiguities          L   XP   W   L       S   M
                ============================   ==   =   =       =   =
   1  CPFLTEISLH LKVHQHPNIA TIHQVLNIED FAIIILMDHF SQGDLFTNII

          S X            X
          = =            =
  51  DRQIFTNNSH RKVPFTDFET QLLMKNAMLQ LIEAIEYCHE NNIVHCDLKP

 101  ENIMVRYNPY VVRPTINNNN NNGEDDLCYA NSIIDYNELH LVLIDFGLAM

 151  DSATICCNEC RGSSFYMAPE RTTNYNTHRL INQLIDMNQY ESIEINGTTV

                                                      F  Y
                                                      =  =
 201  TKSNCKYLPT LAGDIWSLGV LFINITCSRN PWPIASPDNN CNNEVIKNXM

      L     V     L         L          D  D
      =     =     =         =          =  =
 251  MNNNKADLSK IIPISSQFNR ILDRIFKLNP NGRIGL
```

Fig 105

233c gp2

```
   1  NGTKQQITSI IDNLNKADLP KDVEVVICPP ALYLGLAVEQ NKQPTVAIGA

                                                 X
                                                 =
  51  QNVFDKSCGA PTGETCASQI LDVGASWTLT GHSERRTIIK ESDEFIAEKT

                 X           X     X     GCWFQDXX       XGX
                 =           =     =     ========       ===
 101  KFALDTGVKV ILCIGETLEE RKGGVTLDVC ARQLDAVSKI VSDWSNIVVA

                                          missing  sequence
                                          ==========
        LXNWVNFXPX  XKMLKKPTRX*T    X  X XXX
        ==========  ==========      =  = ===
 151  YEPVWAIGTG LAATPEDAEE THKGIRAHLA KSIGAEQAEK TRILYGGSVN

       missing  sequence
      ==========
 201  GKNAKDFKDK ANVDGFLVGG ASLKPEFVDI IKSR
```

Fig 106